

A MULTIPLE REGRESSION MODEL FOR ANGULAR RESPONSES

By

SCOTT P. MORRISON

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1995

UNIVERSITY OF FLORIDA LIBRARIES

To my wife Alicia for her support
and my parents for their motivation.

ACKNOWLEDGEMENTS

I wish to thank Dr. Ramon Littell for being my advisor and supervisor for the last three years at the University of Florida. Without his guidance and suggestions this dissertation could never have been accomplished. His allowance for my annual summer hiatus allowed needed breaks from my studies. He has acted not only as my advisor, but as my friend. I would also like to thank Dr. Brett Presnell for making a simple suggestion from which most of this work was generalized. He has also spent innumerable hours answering my questions, and helping me through stumbling blocks in this research. Additionally, I wish to thank Dr. Jim Booth and Dr. Malay Ghosh for attending numerous meetings and for their willingness to answer questions relating to this research. Finally I wish to thank Dr. John Middlebrooks and Dr. Thomas Walker for being present for my oral exams.

I would like to thank my family, Mom, Dad, Russ and Jackie, for their constant support throughout my schooling. The encouragement that my parents gave me has allowed me to achieve a level of success which I could never have imagined. Numerous other people, including students, peers and other teachers have helped me in ways they may not realize.

Finally, I wish to thank my wife, Alicia, for her constant support. Her love, kindness, smile and humor gave me hope throughout the difficult times. She always seemed to know when to suggest a much needed diversion, as the complications of school were beginning to overburden me. I will always feel lucky to have the love and support of my best friend—my wife.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	viii
ABSTRACT	xi
CHAPTERS	
1 INTRODUCTION	1
2 DIRECTIONAL DATA	4
2.1 Introduction and Sample Statistics	4
2.2 Distributions	7
3 PRIOR WORK	21
3.1 Regression Models	21
3.2 Classification/ANOVA Tests	29
4 A NEW REGRESSION MODEL	35
4.1 Introduction and Motivation	35
4.2 Model Specification	36
4.3 Estimation	42
4.4 Inference	47
5 SIMULATION STUDIES	60
5.1 Introduction	60
5.2 Parameter Estimates	60
5.3 Error Rates of Tests	69
5.4 Circular R^2 statistics	70

6	EXAMPLES	92
6.1	Introduction	92
6.2	Fisher and Lee (1992) Periwinkle Data	92
6.3	Light Intensity and Turtle Hatchling Data	98
6.4	Anderson and Wu (1994) Automotive Flywheel Data	106
6.5	Butterfly Data	109
7	CONCLUSIONS AND FUTURE RESEARCH	119
7.1	Conclusions	119
7.2	Future Research	120
APPENDICES		
A	TRIGONOMETRIC MOMENTS OF THE OFFSET NORMAL DISTRIBUTION	123
B	ESTIMATING THE PARAMETERS OF THE FISHER AND LEE MEAN MODEL	127
C	ESTIMATING THE LENGTH OF A VECTOR GIVEN THE DIRECTION	133
D	NON-NEGATIVE DEFINITENESS OF THE HESSIAN	135
REFERENCES		137
BIOGRAPHICAL SKETCH		140

LIST OF TABLES

<u>Table</u>	<u>Page</u>
2.1 Comparison of method of moments and maximum likelihood estimates of μ for the offset normal distribution.	15
2.2 Comparison of method of moments and maximum likelihood estimates of κ^* for the offset normal distribution.	16
2.3 Estimated relative bias for the maximum likelihood estimators.	17
2.4 Comparison of alternative estimates of κ^*	19
5.1 Comparison of Type I error rates for two asymptotic tests.	71
5.2 Power estimates for two asymptotic tests.	72
6.1 Fisher and Lee (1992) blue periwinkle data.	93
6.2 Analysis of the Fisher and Lee data using the new model.	93
6.3 Sheila Colwell's turtle hatchling data.	101
6.4 Nested ANOVA analysis of the turtle data using our new model. ...	102
6.5 Regression analysis of turtle data using the new model.	105
6.6 Preliminary analysis of the Anderson and Wu (1994) data using the new model.	108
6.7 Ordering of factors.	109
6.8 Factor comparisons for the final model.	110
6.9 Estimated mean directions and concentrations for the final model. ..	111
6.10 Analysis of the Long-Tailed Skipper data.	115
6.11 Analysis of the Gulf Fritillary data.	115

6.12 Analysis of the Cloudless Sulphur data.	117
B.1 Simulation results of Price's Algorithm: two dimensional case.	130
B.2 Simulation results of Price's Algorithm: three dimensional case.	132

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2.1 A directional observation.	4
2.2 Comparison of the concentration functions for the von Mises and offset normal distributions.	12
2.3 Comparisons of density functions for some choices of κ , von Mises - -, offset normal —.	13
3.1 A simple example showing the periodicity of the Gould model.	23
3.2 Log-likelihood component for the Fisher and Lee model with the periwinkle data.	26
3.3 Log-likelihood component of the Fisher and Lee model for a simulated data set with $n = 10$, $\kappa = 1.0$, $\beta_1 = .1$ and $\beta_2 = .1$. The global maximum is at *.	27
4.1 Actual data in two-space, where only direction is recorded.	36
4.2 Fitting of the mean in two-space, and its associated model of the mean direction.	37
4.3 Possible curves for the mean direction, by allowing quadratic terms in the model.	41
4.4 Implication of projecting a confidence region in two-space onto the unit circle.	54
5.1 Estimation of the parameter β_{12} for sample size 20.	63
5.2 Estimation of the parameter β_{12} for sample size 50.	64
5.3 Estimation of the parameter β_{12} for sample size 100.	65
5.4 Estimation of the parameter β_{02} for sample size 20.	66
5.5 Estimation of the parameter β_{12} for sample size 50.	67

5.6	Estimation of the parameter β_{12} for sample size 100.	68
5.7	Empirical distribution of the circular R^2 values, $n=20$	74
5.8	Empirical distribution of the circular R^2 values, $n=20$	75
5.9	Empirical distribution of the circular R^2 values, $n=20$	76
5.10	Empirical distribution of the circular R^2 values, $n=20$	77
5.11	Empirical distribution of the circular R^2 values, $n=20$	78
5.12	Empirical distribution of the circular R^2 values, $n=20$	79
5.13	Empirical distribution of the circular R^2 values, $n=50$	80
5.14	Empirical distribution of the circular R^2 values, $n=50$	81
5.15	Empirical distribution of the circular R^2 values, $n=50$	82
5.16	Empirical distribution of the circular R^2 values, $n=50$	83
5.17	Empirical distribution of the circular R^2 values, $n=50$	84
5.18	Empirical distribution of the circular R^2 values, $n=50$	85
5.19	Empirical distribution of the circular R^2 values, $n=100$	86
5.20	Empirical distribution of the circular R^2 values, $n=100$	87
5.21	Empirical distribution of the circular R^2 values, $n=100$	88
5.22	Empirical distribution of the circular R^2 values, $n=100$	89
5.23	Empirical distribution of the circular R^2 values, $n=100$	90
5.24	Empirical distribution of the circular R^2 values, $n=100$	91
6.1	Fisher and Lee (1992) data.	94
6.2	Comparison of three models.	96
6.3	Approximate 95% pointwise confidence bands for the mean direction.	97
6.4	Comparison of two approximate 95% prediction bands for a future observation.	99
6.5	Sheila Colwell's turtle hatchling data.	100

6.6	Fitted values and prediction bands for the turtle data with a nested design.	103
6.7	Plotted light intensities versus angle for location 13 on 9/1.	104
6.8	Confidence and prediction bands for the turtle data.	106
6.9	Plot of the Long Tailed Skipper data.	112
6.10	Plot of the Gulf Fritillary data.	113
6.11	Plot of the Cloudless Sulpher data.	114
6.12	Fitted model along with 95% confidence and prediction bands for the Long-Tailed Skipper data.	116
6.13	Fitted model along with 95% confidence and prediction bands for the Gulf Fritillary data.	117
6.14	Fitted model along with 95% confidence and prediction bands for the Cloudless Sulphur data.	118

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

A MULTIPLE REGRESSION MODEL FOR ANGULAR RESPONSES

By

Scott P. Morrison

December 1995

Chairman: Ramon C. Littell

Major department: Statistics

The multiple linear regression model allows for modeling a set of scalar responses as functions of any number of covariates. These covariates can be dummy variables, representing the presence or absence of a factor, and/or continuous variables. When the response is not a scalar variable, but a direction in two dimensions, standard linear statistics can not be used. Attempts to relate these angular responses to a set of covariates via regression models, similar to those used for scalar responses, have not been totally successful. Estimation difficulties often arise with these models, due to identifiability problems.

This dissertation proposes a single useful model that does not suffer from these estimation difficulties. This new model is based on a little-used distribution in the directional data literature, the offset normal. Distributional properties of the offset normal are described and the offset normal is compared to the von Mises distribution, the most commonly used directional distribution. The new model is then introduced along with a simple iterative estimation procedure. A number of inferential procedures based on standard asymptotic results are also discussed. Small simulation studies are

used to support the inferential procedures, and a number of examples are described to show the variety of situations in which this new model can be applied.

CHAPTER 1 INTRODUCTION

The classical linear model is used to characterize a simple relationship between a scalar response variable and a set of covariates. In many settings the covariates are scalar variables themselves and are referred to as regressor variables. The covariates can also be dummy variables, representing the presence or absence of a factor (parameter) in the model. The resulting model is often referred to as a classification model or analysis of variance (ANOVA) model. Models in which both types of covariates are included are possible for the classical linear model.

In certain research settings, the response is not a scalar variable, but a direction in two dimensions requiring regression models for a directional response. Fisher and Lee (1992) discussed modeling the direction of blue periwinkles as a function of distance moved. Anderson and Wu (1994a, 1994b) discussed an industrial experiment involving a 2^4 factorial design where the response is the directional adjustment necessary for balancing an automotive flywheel. Walker and Littell (1994) described an observational study of the directions that butterflies travel, where the explanatory variables are numerous, including distance from the coast, percentage cloud cover, and time of day.

Models proposed in the literature for these settings are not completely acceptable for two reasons. First, when the response is a direction, the formulation of models for multiple linear regression settings and analysis of variance settings have been treated as separate problems. In particular, models proposed by Fisher and Lee and Gould (1969) were for the continuous covariate case. Using these models with

dummy variables reveal inherent difficulties in their formulation. Tests proposed for the classification variable case are not based on parametric models, and generalizing them to include continuous covariates is impossible. Second, estimation difficulties often exist for the multiple regression models in the literature. In the cases of the Fisher and Lee and Gould models the log-likelihood functions, which are maximized to yield maximum likelihood estimates, are multimodal and graphical exploration of the likelihood is necessary to aid in estimation. Graphical exploration becomes increasingly more difficult as the number of covariates increases.

This dissertation proposes a single model that is applicable for all types of covariates, resulting in a model along the lines of the classical linear model for scalar responses. The model necessitates that both the mean direction and concentration (variance) of a directional observation are functions of the covariate vector.

Chapter 2 begins with some introductory notation for directional data. A seldom used distribution, the offset normal, is introduced and simple distributional properties are revealed. The distribution is also compared to the standard distribution for directional data, the von Mises, and an interesting relationship between the two distributions is discussed. A small simulation study investigates simple parameter estimation for the offset normal distribution in the independent, identically distributed (iid) case, and compares the bias of these estimates to the bias in using the von Mises distribution. Chapter 3 describes more fully the previously suggested regression models and their flaws. Three classification, or ANOVA, techniques are also described and compared. Chapter 4 introduces the new regression model and describes its formulation. The EM algorithm is used to find maximum likelihood estimates of the parameters and their uniqueness is guaranteed. Several inferential procedures, including techniques for model specification and computing confidence and prediction intervals, are described based on large sample asymptotics. Finally, a generalization for the classification setting is described that is related to method of

moments estimation. Chapter 5 gives the results of a small simulation study. The goal of the simulation study is to test the appropriateness of several approximations and to test distributional properties of parameters and goodness of fit statistics. In Chapter 6 the new model is fitted to a number of data sets that show its wide range of applicability. Some conclusions and recommendations for further research are included in Chapter 7.

CHAPTER 2 DIRECTIONAL DATA

2.1 Introduction and Sample Statistics

Directional data in two dimensions arise quite commonly in geology, biology and many ecological settings. A directional observation is typically referred to by its associated angle, θ , from some chosen zero direction. The observation can also be viewed as a point on the unit circle, $(\cos(\theta), \sin(\theta))$, as Figure 2.1 shows.

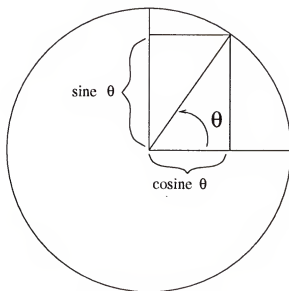


Figure 2.1. A directional observation.

Throughout, it will be assumed that $\theta_i \in [0, 2\pi)$, unless otherwise noted. In general, standard statistical techniques for data on the real line cannot be used to analyze sets of directional observations. The simplest example of an error being found by applying scalar statistics to directional data is by finding the mean of two angles

lying 2δ radians apart. If δ is very small and the points are $2\pi - \delta$ and $0 + \delta$ (0 and 2π are indistinguishable on the circle), then their arithmetic mean is π , even though both angles are very close to zero. If the data were rotated slightly (or the origin changed), the arithmetic mean would change dramatically.

A more appropriate measure of mean direction for directional observations is the sample mean direction, $\bar{\theta}$, defined as the solution to

$$\sum_{i=1}^n \sin(\theta_i - \bar{\theta}) = 0. \quad (2.1)$$

By defining

$$S = \sum_{i=1}^n \sin(\theta_i), \quad C = \sum_{i=1}^n \cos(\theta_i),$$

and requiring $\bar{\theta} \in [0, 2\pi)$, equation (2.1) implies

$$\bar{\theta} = \begin{cases} \tan^{-1}\left(\frac{S}{C}\right) & S \geq 0, C > 0 \\ \tan^{-1}\left(\frac{S}{C}\right) + \pi & C < 0 \\ \tan^{-1}\left(\frac{S}{C}\right) + 2\pi & S < 0, C > 0 \\ \frac{\pi}{2} & S > 0, C = 0 \\ \frac{3\pi}{2} & S < 0, C = 0 \\ \text{undefined} & S = 0, C = 0 \end{cases} \quad (2.2)$$

The sample mean direction, $\bar{\theta}$, represents the direction of the point (C, S) in the two dimensional plane. For simplicity the term $\tan^{-1}(a, b)$ will be used throughout the rest of this dissertation to represent the direction of the point (a, b) in the two dimensional plane, i.e.

$$\tan^{-1}(a, b) = \begin{cases} \tan^{-1}\left(\frac{b}{a}\right) & b \geq 0, a > 0 \\ \tan^{-1}\left(\frac{b}{a}\right) + \pi & a < 0 \\ \tan^{-1}\left(\frac{b}{a}\right) + 2\pi & b < 0, a > 0 \\ \frac{\pi}{2} & b > 0, a = 0 \\ \frac{3\pi}{2} & b < 0, a = 0 \\ \text{undefined} & b = 0, a = 0 \end{cases}, \quad (2.3)$$

implying

$$\bar{\theta} = \tan^{-1}(C, S).$$

The resultant length, R , defined as

$$R = (C^2 + S^2)^{\frac{1}{2}} = \sum_{i=1}^n \cos(\theta_i - \bar{\theta})$$

is a sample statistic that gives information about the variability in the data. R will be large and close to n when the data are concentrated and approaches zero as the data become more dispersed. Hence, an appropriate measure of variability for directional observations is the sample circular variance (Mardia, 1972), S_0 , defined as

$$S_0 = 1 - \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \bar{\theta}) = \frac{n - R}{n}. \quad (2.4)$$

Since R is bounded from above by n , $S_0 \in [0, 1]$ and behaves similarly to the variance for linear data. It is small when there is little variability in the data (highly concentrated) and large when the data are highly variable (not concentrated). A useful function of the circular variance, defined on the positive real line, is the squared circular standard deviation, s_0^2 , defined as

$$s_0^2 = -2 \log(1 - S_0). \quad (2.5)$$

This transformation has some theoretical justification. Equating the first trigonometric moments (the appropriate moments for directional data) of the von Mises distribution and the wrapped normal distribution, both of which will be discussed in the next section, gives the above identity. (See Mardia (p. 24, 66) for a thorough discussion.) These measures of concentration are not the only measures generally accepted, but they will be sufficient for this dissertation.

As with linear data it is often sufficient to describe a set of data by only a few sample statistics; however it is usually more informative to assume a particular form for the underlying distribution.

2.2 Distributions

Throughout the rest of this dissertation it will be assumed that the directional observations, θ_i , $i = 1, \dots, n$, follow some underlying distribution with circular mean directions μ_i and concentration parameters κ_i . Concentration is the standard term used to explain variability in directional data. (Generally the concentration parameter behaves like the inverse of σ^2 .) Later certain parametric forms will be assumed for these population parameters.

2.2.1 von Mises Distribution

The von Mises distribution (also referred to as the Fisher and Langevin distributions for two dimensions), denoted throughout as vM, is the distribution most commonly used for inference in modeling directional data. This can be attributed to its possession of many, but not all, properties on the circle similar to the normal distribution on the real line. (Together with the wrapped normal distribution, defined as a normal random variable modulo 2π , these two distributions jointly possess all the attractive properties of the normal distribution on the real line.) The form of the von Mises density is

$$f_{\text{vM}}(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp(\kappa \cos(\theta - \mu)), \quad 0 < \theta, \mu \leq 2\pi, \text{ and } \kappa \geq 0, \quad (2.6)$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind and order zero and has the infinite series representation

$$I_0(\kappa) = \sum_{r=0}^{\infty} \frac{\kappa^{2r}}{2^{2r}(r!)^2}.$$

The parameter μ is defined to be the circular mean direction and κ is a concentration parameter.

The von Mises distribution is symmetric and unimodal and generally has a bell-shape similar to the normal distribution. For $\kappa = 0$ the distribution corresponds to the uniform distribution on the circle, and as κ approaches infinity the distribution approaches a point distribution at μ . For κ large (typically ≥ 2), the von Mises distribution can be approximated by a normal distribution, with mean μ and variance $\frac{1}{\kappa}$. Taking a second order Taylor series approximation of $\cos(\theta - \mu)$ about zero, gives

$$\cos(\theta - \mu) \approx 1 - \frac{(\theta - \mu)^2}{2}.$$

For large κ , Abramowitz and Stegun (1970) and Mardia (1972) give the result

$$I_0(\kappa) \approx \frac{\exp(\kappa)}{(2\pi\kappa)^{\frac{1}{2}}}.$$

Inserting these two approximations into equation (2.6) gives the desired result.

For a random sample of n observations, all having the same mean direction and concentration, it is easy to verify that the maximum likelihood estimates of μ and κ satisfy

$$\hat{\mu} = \bar{\theta} \quad \text{and} \quad A(\hat{\kappa}) = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \bar{\theta}) = \bar{R},$$

where $A(\cdot) = I_1(\cdot)/I_0(\cdot)$ is a ratio of first order Bessel functions of order one and zero respectively. The $A(\cdot)$ function does not have a closed form inverse, so some approximation, computation, or use of tables is necessary for finding $\hat{\kappa}$.

Before describing the method of moments estimates for the von Mises distribution, a general result is required. It can be shown that, for any symmetric (about μ) and unimodal directional distribution, the first trigonometric moments are

$$\mathbf{E}(\cos \theta) = \rho \cos \mu \tag{2.7}$$

and

$$\mathbf{E}(\sin \theta) = \rho \sin \mu. \tag{2.8}$$

The parameter ρ is a concentration parameter and is equal to $A(\kappa)$ for the von Mises distribution. Thus, the method of moments estimators for the von Mises distribution result in the equations

$$\hat{\rho} = A(\hat{\kappa}) = \bar{R} \quad \text{and} \quad \hat{\mu} = \bar{\theta}.$$

Hence, for the von Mises distribution, the maximum likelihood and the method of moments estimators are identical.

2.2.2 Offset Normal Distribution

A separate, but related, distribution is the offset normal distribution, denoted henceforth as ON. The particular form of the density is

$$f_{\text{ON}}(\theta) = (2\pi)^{-1} \exp\left(-\frac{\kappa^{*2}}{2}\right) \left[1 + \sqrt{2\pi}\kappa^* \cos(\theta - \mu) \times \exp\left(\frac{\kappa^{*2} \cos^2(\theta - \mu)}{2}\right) \Phi(\kappa^* \cos(\theta - \mu))\right],$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. (Mardia (p. 52) defines the distribution in a slightly more general form.) The density can be shown to be positive by an application of Mill's ratio (Resneck, 1992). Like the von Mises distribution, this distribution is symmetric (about μ), unimodal, and has a bell shape. For $\kappa^* = 0$, the distribution corresponds to the uniform distribution, and as κ^* approaches infinity the distribution approaches a point distribution at μ , again like the von Mises distribution.

Unfortunately, very few, if any, applications of this distribution have been discussed in the framework of directional data. This can be attributed to the complicated form of the offset normal distribution. Mardia (p. 53) states that “the distribution theory necessary for statistical inference (based on this distribution) is formidable.” The distribution has been used in a nonparametric setting involving the multivariate

sign test and first appeared in Klotz (1964). However, simple properties, including moments, seem to be missing from the literature. Appendix A gives the details on finding the trigonometric moments. For parameter estimation, only the first trigonometric moments are necessary and take the form:

$$E(\cos \theta) = \rho \cos \mu \quad (2.9)$$

and

$$E(\sin \theta) = \rho \sin \mu, \quad (2.10)$$

where $\rho = S(\kappa^*)$, and

$$S(\kappa^*) = \frac{1}{4} \sqrt{2\pi} \exp\left(-\frac{1}{2}\kappa^{*2}\right) \sum_{i=0}^{\infty} \frac{\kappa^{*(2i+1)}(2i+1)!}{2^{2i}(i!)^2(i+1)!}. \quad (2.11)$$

The method of moment estimators for the offset normal distribution satisfy

$$\hat{\rho} = S(\hat{\kappa}^*) = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i - \bar{\theta}) = \bar{R} \quad \text{and} \quad \hat{\mu} = \bar{\theta}.$$

As with the von Mises distribution, computation or approximation is necessary to find κ^* as the function $S(\cdot)$ is not invertible in closed form. Maximum likelihood estimators are more difficult to find, since the scoring equations do not have closed form. A technique to obtain maximum likelihood estimates will be presented in Chapter 4. It can be shown that the maximum likelihood estimate and the method of moments estimators for κ^* are not equivalent.

2.2.3 Generation of the Distributions

The two aforementioned distributions are related to each other through their connection with the bivariate normal distribution. The von Mises distribution can be derived as the conditional distribution of the direction, from the origin, of a bivariate normal random vector, with independent components having equal variances, given

that the length of the vector is fixed. The offset normal distribution can be obtained as the marginal distribution of the direction, from the origin, of a bivariate normal random vector, with independent components having equal variances. Specifically, assume that

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim BN \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, I \right), \quad (2.12)$$

where the mean vector (μ_1, μ_2) has polar representation $(\kappa \cos \mu, \kappa \sin \mu)$. The conditional distribution of the direction of (y_1, y_2) , conditioned on the length of the vector being equal to one, can be found by transforming (y_1, y_2) to its polar coordinates $(L \cos \theta, L \sin \theta)$. The resulting density is von Mises with mean direction μ and concentration parameter κ , i.e.

$$(\theta|L=1) \sim \text{vM}(\mu, \kappa).$$

Similarly, using equation (2.12) and the polar representation of (y_1, y_2) , the marginal distribution of the direction of (y_1, y_2) is offset normal with mean direction μ and concentration parameter κ , i.e.

$$\theta \sim \text{ON}(\mu, \kappa).$$

Thus, it can be seen that both the von Mises distribution and the offset normal distribution can be generated from a bivariate normal distribution with identity covariance matrix. The two distributions are equivalent if and only if the direction, θ , is independent of the length of the vector L . This is true only if $\kappa = 0$, which results in θ being uniformly distributed over $[0, 2\pi)$.

2.2.4 Comparison of the Distributions

Besides being related to one another through their generation from the bivariate normal distribution, the two families of distributions can approximate one another

for suitably chosen concentration parameters. This can most easily be illustrated by equating their first trigonometric moments. In Figure 2.2 the two functions $A(\cdot)$ and $S(\cdot)$ are plotted over the same values of κ .

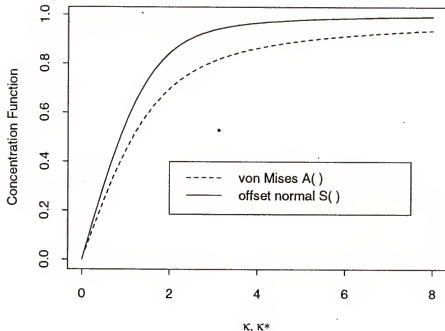


Figure 2.2. Comparison of the concentration functions for the von Mises and offset normal distributions.

To equate the two distributions values of κ and κ^* such that $A(\kappa) = S(\kappa^*)$, which correspond to the first centered trigonometric moments, are needed. Appropriate comparisons for the two distributions would use the equivalent concentration values and assume equal mean directions. Figure 2.3 shows plots comparing the von Mises and offset normal densities where κ was chosen as .1, .5, 1, 2, 2.5, and 5, and the appropriate values of κ^* (.0798, .3945, .7643, 1.3742, 1.6110, and 2.3790), which correspond to $(S^{-1}(A(.1)), \dots, S^{-1}(A(2.5)))$. The mean directions are taken to be π so that the plots are symmetric over the interval $[0, 2\pi)$.

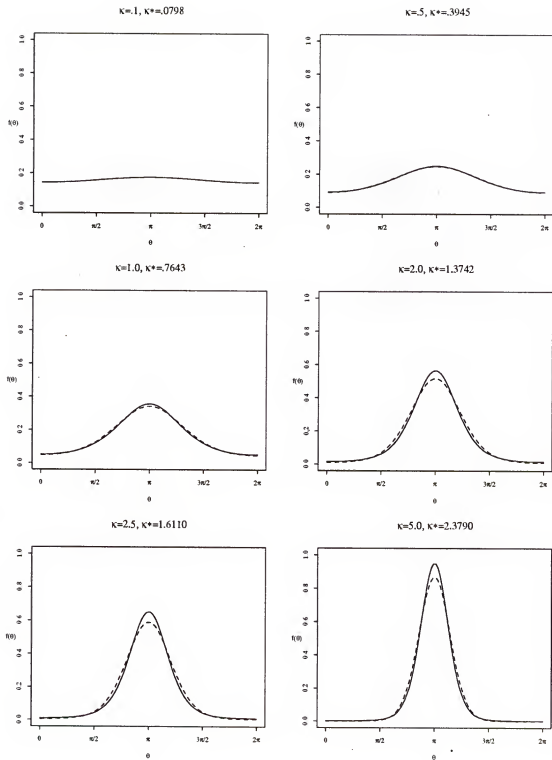


Figure 2.3. Comparisons of density functions for some choices of κ , von Mises - -, offset normal —.

The plots reveal that the distributions are almost indistinguishable for small and large values of $\kappa(\kappa^*)$. For intermediate values of $\kappa(1.0 - 2.5)$ the two distributions are slightly different, with the offset normal distribution having the heavier tails and higher peak. However, for a real data setting, it would seem difficult to distinguish between the two densities.

2.2.5 Simulation Study of Parameter Estimation

It is well known that the maximum likelihood (hence, method of moments) estimator of κ for the von Mises distribution is heavily biased for small samples and/or data that are not extremely concentrated. (See Fisher, 1993 p. 88, for a discussion and some recommended adjustments and Best and Fisher, 1981, for results of a simulation study.) Since the two distributions have such similar shapes, it is of interest to find whether the estimation of κ^* for the offset normal distribution also has these complications.

A small simulation study was designed to answer this question. The format of this simulation mirrors that of Best and Fisher. Samples of pseudo-random variables from the offset normal distribution with mean zero were generated, of size $n = 10, 20$ and 100 , by generating pairs of independent normal random variables using the Ziggurat method described in Marsaglia and Tsang (1984) and extracting only the direction of these pairs. The concentration parameters for the offset normal distribution were chosen as $\kappa^* = .0798, .3945, .7643, 1.3742, 1.6110$ and 2.3790 . (These values were taken so that the associated κ values for the von Mises were simple, known and equivalent to those used by Best and Fisher.) Ten thousand (10,000) data sets were generated for each combination of sample size and concentration, then both method of moments and maximum likelihood estimates were found for μ and κ^* of each data set. (The technique for finding maximum likelihood estimates will be discussed in Chapter 4.)

Table 2.1 gives the estimated Monte Carlo directional means, denoted by $E(\hat{\mu}_{ml})$ and $E(\hat{\mu}_{mm})$, of the maximum likelihood and method of moments estimators of $\mu = 0$. The true mean direction, μ , was not changed because rotation does not affect the model.

Table 2.1. Comparison of method of moments and maximum likelihood estimates of μ for the offset normal distribution.

n	$\kappa^*(\kappa)$	$E(\hat{\mu}_{ml})$	$E(\hat{\mu}_{mm})$
10	.0798(0.1)	-0.02507	-0.02838
	.3945(0.5)	-0.01898	-0.01828
	.7643(1.0)	-0.00564	-0.00629
	1.3742(2.0)	0.00060	-0.00022
	1.6110(2.5)	-0.00197	-0.00204
	2.3790(5.0)	0.00039	0.00014
20	.0798(0.1)	-0.00528	-0.00691
	.3945(0.5)	-0.00937	-0.00943
	.7643(1.0)	0.00487	0.00476
	1.3742(2.0)	0.00077	0.00132
	1.6110(2.5)	-0.00118	-0.00109
	2.3790(5.0)	0.00023	0.00010
100	.0798(0.1)	0.01604	0.01640
	.3945(0.5)	0.00265	0.00295
	.7643(1.0)	0.00077	0.00083
	1.3742(2.0)	0.00019	0.00027
	1.6110(2.5)	-0.00033	-0.00029
	2.3790(5.0)	0.00011	0.00002

Table 2.1 shows that both the method of moments and maximum likelihood estimators of μ are approximately unbiased. The two estimators are also almost indistinguishable from one another. This is not surprising since the offset normal distribution matches the von Mises distribution so well and the two estimators are identical for the von Mises distribution.

Table 2.2 gives the Monte Carlo mean, denoted by $E(\hat{\kappa}_{ml}^*)$, of $\hat{\kappa}_{ml}^*$, the maximum likelihood estimator of κ^* . It also gives the Monte Carlo medians, denoted by $M(\hat{\kappa}_{ml}^*)$ and $M(\hat{\kappa}_{mm}^*)$, of the maximum likelihood and method of moments estimators of κ^* .

Table 2.2. Comparison of method of moments and maximum likelihood estimates of κ^* for the offset normal distribution.

n	$\kappa^*(\kappa)$	$E(\hat{\kappa}_{ml}^*)$	$M(\hat{\kappa}_{ml}^*)$	$M(\hat{\kappa}_{mm}^*)$
10	.0798(0.1)	.4805	.4474	.4453
	.3945(0.5)	.6153	.5726	.5740
	.7643(1.0)	.9190	.8881	.8865
	1.3742(2.0)	1.5594	1.5018	1.5141
	1.6110(2.5)	1.8102	1.7420	1.7664
	2.3790(5.0)	2.6909	2.5677	2.6386
20	.0798(0.1)	.3361	.3142	.3140
	.3945(0.5)	.4996	.4829	.4825
	.7643(1.0)	.8381	.8266	.8254
	1.3742(2.0)	1.4539	1.4306	1.4341
	1.6110(2.5)	1.7025	1.6745	1.6852
	2.3790(5.0)	2.5259	2.4773	2.5106
100	.0798(0.1)	.1595	.1504	.1500
	.3945(0.5)	.4134	.4128	.4127
	.7643(1.0)	.7790	.7776	.7771
	1.3742(2.0)	1.3890	1.3847	1.3870
	1.6110(2.5)	1.6296	1.6243	1.6255
	2.3790(5.0)	2.4083	2.3995	2.4052

It can be noted from Table 2.2 that both the method of moments and maximum likelihood estimates of κ^* are biased upwardly. This result is very similar to that shown by Best and Fisher for the von Mises distribution. It can also be seen that the method of moments and maximum likelihood estimators of κ^* are very similar, although not identical. Table 2.3 gives the estimated relative bias for the maximum likelihood estimates of κ^* . The estimated relative bias of \hat{x} in estimating x is defined

as

$$\frac{\hat{x} - x}{x}.$$

In parentheses next to each estimate is the estimated relative bias for the estimation of the associated κ for the von Mises distribution found by Best and Fisher. It can be seen that the order of magnitude of the relative bias for estimation of κ^* for the offset normal distribution is on the same order of magnitude for small values of κ as estimation of the relative bias of κ for the von Mises distribution. For large values of κ the relative bias is smaller for the offset normal distribution.

Table 2.3. Estimated relative bias for the maximum likelihood estimators.

n	$\kappa^*(\kappa)$	$E(\hat{\kappa}_{\text{ml}}^*)$	$M(\hat{\kappa}_{\text{ml}}^*)$	$M(\hat{\kappa}_{\text{mun}}^*)$
10	.0798(0.1)	5.0213(5.3)	4.6065(4.6)	4.5802
	.3945(0.5)	.5597(.64)	.4515(.48)	.4550
	.7643(1.0)	.2024(.29)	.1620(.17)	.1599
	1.3742(2.0)	.1348	.0929	.1018
	1.6110(2.5)	.1236(.34)	.0813(.15)	.0965
	2.3790(5.0)	.1311(.41)	.0793(.15)	.1091
20	.0798(0.1)	3.2118(3.3)	2.9373(2.9)	2.9348
	.3945(0.5)	.2664(.26)	.2241(.22)	.2231
	.7643(1.0)	.0966(.12)	.0815(.09)	.0799
	1.3742(2.0)	.0580	.0410	.0436
	1.6110(2.5)	.0568(.14)	.0394(.07)	.0461
	2.3790(5.0)	.0617(.15)	.0413(.09)	.0553
100	.0798(0.1)	.9987(1.0)	.8847(.9)	.8797
	.3945(0.5)	.0479(.06)	.0464(.04)	.0461
	.7643(1.0)	.0192(.02)	.0174(.01)	.0167
	1.3742(2.0)	.0108	.0076	.0093
	1.6110(2.5)	.0115(.02)	.0083(.01)	.0090
	2.3790(5.0)	.0123(.02)	.0086(.01)	.0110

2.2.6 Possible Improved Estimators

Based on the observed bias for the concentration parameters of the offset normal distribution, four alternative estimation techniques are suggested. Using the ideas of Best and Fisher and some experimentation the first alternative estimator is of the form

$$\hat{\kappa}_1^* = \begin{cases} \max \{ \hat{\kappa}^* - (-53.2278 + 20.4563/\hat{\kappa}^* + 44.8231\hat{\kappa}^*)/n^{1.5}, 0 \} & \hat{\kappa}^* < 1.0 \\ \max \{ \hat{\kappa}^* - (-1.4185 + 1.3425/\hat{\kappa}^* + 1.4638\hat{\kappa}^*)/n, 0 \} & \hat{\kappa}^* \geq 1.0 \end{cases} \quad (2.13)$$

The particular form of $\hat{\kappa}_1^*$ was found by plotting the estimated value of $\hat{\kappa}^*$ versus the estimated bias. It was found that there were two distinct ranges of $\hat{\kappa}^*$ where the bias was on different orders. Over these separate ranges the alternative estimator was estimated by least squares.

The second alternative estimator is the jackknife estimator, defined as

$$\hat{\kappa}_2^* = \max \left\{ n\hat{\kappa}^* - \frac{n-1}{n} \sum_{i=1}^n \hat{\kappa}_{-i}^*, 0 \right\}, \quad (2.14)$$

where $\hat{\kappa}_{-i}^*$ is the estimate of κ^* when the i^{th} observation is removed from the sample.

The last two alternative estimators are bootstrap estimators. The first is a non-parametric bootstrap estimator of the form

$$\hat{\kappa}_3^* = \max \{ 2\hat{\kappa}^* - \bar{\kappa}_{np}^*, 0 \}, \quad (2.15)$$

where $\bar{\kappa}_{np}^*$ is the sample mean for estimates of κ^* over random resamples taken, with replacement, from the original data. The parametric bootstrap estimator has the same form as that in equation (2.15) with $\bar{\kappa}_{np}^*$ replaced by $\bar{\kappa}_p^*$, i.e.

$$\hat{\kappa}_4^* = \max \{ 2\hat{\kappa}^* - \bar{\kappa}_p^*, 0 \}. \quad (2.16)$$

In this case $\bar{\kappa}_p^*$ is the sample mean of estimates of κ^* over random samples of pseudo-random variables generated from the offset normal with parameters $(\mu, \kappa^*) = (\hat{\mu}, \hat{\kappa}^*)$.

Table 2.4 gives the means of the alternative estimators based on one thousand replications. One thousand replicates was chosen due to the large amount of time necessary for the bootstrap estimators. The bootstrap estimators are based on four hundred resamples for each replicate.

Table 2.4. Comparison of alternative estimates of κ^* .

n	$\kappa^*(\kappa)$	$E(\hat{\kappa}_1^*)$	$E(\hat{\kappa}_2^*)$	$E(\hat{\kappa}_3^*)$	$E(\hat{\kappa}_4^*)$
10	.0798(0.1)	0.2213	0.3112	0.2894	0.3012
	.3945(0.5)	0.3577	0.4438	0.4262	0.4383
	.7643(1.0)	0.6858	0.7525	0.7363	0.7504
	1.3742(2.0)	1.3571	1.3356	1.3259	1.3462
	1.6110(2.5)	1.6026	1.5595	1.5507	1.5755
	2.3790(5.0)	2.3764	2.2631	2.2821	2.3044
20	.0798(0.1)	0.1645	0.2261	0.2161	0.2194
	.3945(0.5)	0.3405	0.3913	0.3828	0.3866
	.7643(1.0)	0.7250	0.7475	0.7431	0.7463
	1.3742(2.0)	1.3718	1.3704	1.3691	1.3715
	1.6110(2.5)	1.6064	1.6017	1.6017	1.6013
	2.3790(5.0)	2.3943	2.3759	2.3811	2.3807
100	.0798(0.1)	0.1019	0.1209	0.1177	0.1179
	.3945(0.5)	0.3934	0.3908	0.3897	0.3901
	.7643(1.0)	0.7734	0.7672	0.7671	0.7669
	1.3742(2.0)	1.3792	1.3783	1.3787	1.3781
	1.6110(2.5)	1.6132	1.6123	1.6123	1.6113
	2.3790(5.0)	2.3917	2.3917	2.3919	2.3909

As expected, the bias of each alternative estimator is smaller than for either the maximum likelihood or method of moments estimators. The first alternative estimator seems to be the least biased for very small concentrations, although all four estimators show significant bias for extremely small κ^* . This is not surprising, since the distribution is almost indistinguishable from the uniform when κ^* is extremely

small. For large concentrations all estimators seem to be approximately unbiased. This is especially true for large sample sizes. The bootstrap estimators are costly in terms of computing time, due to the iterative nature of the estimates themselves which will be discussed in Chapter 5. The first alternative estimator can perhaps be improved through further simulation studies. The two equations that make up the first alternative estimator could be improved through further simulation studies involving additional values of κ^* . Further simulations should also be done to test whether the improved (alternative) estimating techniques actually increase the amount of variability.

CHAPTER 3 PRIOR WORK

3.1 Regression Models

Throughout this dissertation, it will be assumed that the observed directional observations $(\theta_1, \dots, \theta_n)$ have associated covariate vectors $(\mathbf{x}'_1, \dots, \mathbf{x}'_n)$. The directional observations are assumed to have mean directions (μ_1, \dots, μ_n) and concentration parameters $(\kappa_1, \dots, \kappa_n)$.

A regression model for a directional response is difficult to visualize. For a single covariate the pairs $((\theta_1, x_1), \dots, (\theta_n, x_n))$ can be viewed as points on the surface of a cylinder with unit radius and axis of the cylinder corresponding to the axis of X . Fitting a regression model to such data corresponds to fitting a curve on the surface of this cylinder that attempts to explain the pattern of points on the cylinder. For multiple covariates, the data can be imagined as lying on the surface of a higher dimensional cylinder.

3.1.1 Gould (1969)

Gould proposed a regression model for a directional response that had the form of the multiple linear regression model for a linear response, except the multiple linear regression equation was taken modulo 2π to map it onto the unit circle. In particular, Gould assumed that the directional responses were distributed as von Mises with constant concentration parameter, κ , and mean directions

$$\mu_i = (\mathbf{x}'_i \beta) \bmod 2\pi, \quad (3.1)$$

where \mathbf{x}_i is a vector of covariates and $\boldsymbol{\beta}$ is a vector of unknown "regression" parameters. This model is often referred to as a Barber's Pole model since, for a single covariate, traveling from $-\infty$ to ∞ , the mean direction travels around an infinite cylinder infinitely many times. Gould also proposed a related model for the spherical regression setting based on the Fisher distribution in three dimensions.

Gould suggested an iterative Newton-Raphson algorithm for maximum likelihood parameter estimation. Unfortunately, the associated log-likelihood function for the Gould model is multimodal. Gould (p. 686) recognized the complication of local maxima over the log-likelihood surface; however it was believed that "... as the sample size and the number of distinct concomitant(covariate) vectors in the sample increase, the chance of getting ... (an undesired estimate) ... diminishes, and vanishes in the limit." This can, however, be disproved.

Johnson and Wehrly (1978) pointed out that, for a single covariate and equally spaced design points, the likelihood function is periodic, so that there are an infinite number of parameter values for which the likelihood is maximized. Figure 3.1 shows a plot of this phenomenon. Both sets of curves go perfectly through the two points on the cylinder, so that both curves are defined by different maximum likelihood estimates. Additional curves can be drawn that also go through both points and travel around the cylinder additional times between the two points. This same phenomenon can occur for larger data sets.

The argument of Johnson and Wehrly can be generalized to settings in which the covariate values lie on a lattice. Data on a lattice correspond to data where the distances, between all pairs of points (in each direction) are proportional to one another. The resulting period of the log-likelihood is proportional to the span of the lattice. Equally spaced data provide one, but not the only example of such data.

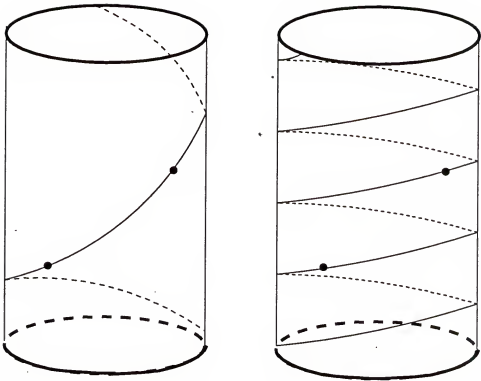


Figure 3.1. A simple example showing the periodicity of the Gould model.

Assuming that the covariates can be placed on a lattice is tantamount to claiming that

$$x_{ij} = A_{ij}\delta_j,$$

where A_{ij} is an integer and δ_j is the width between lattice points in the dimension corresponding to X_j . If $\hat{\beta}_j$ is the maximum likelihood estimator of β_j then

$$\hat{\beta}_j^k = \hat{\beta}_j + 2\pi k/\delta_j$$

is also a maximum likelihood estimator for β_j for all integers k . This can be shown by realizing that the likelihood depends on β only through the log-likelihood component, which for the Gould model corresponds to

$$\sum_{i=1}^n \cos \left(\theta_i - \sum_j \beta_j x_{ij} \right).$$

Now

$$\begin{aligned}
 \sum_{i=1}^n \cos \left(\theta_i - \sum_j \hat{\beta}_j x_{ij} \right) &= \sum_{i=1}^n \cos \left(\theta_i - \sum_j \hat{\beta}_j (A_{ij} \delta_j) \right) \\
 &= \sum_{i=1}^n \cos \left(\theta_i - \sum_j \left(\hat{\beta}_j^k - \frac{2\pi k}{\delta_j} \right) (A_{ij} \delta_j) \right) \\
 &= \sum_{i=1}^n \cos \left(\theta_i - \sum_j (\hat{\beta}_j^k (A_{ij} \delta_j) - 2\pi k (A_{ij})) \right).
 \end{aligned}$$

Since cosine is periodic with period 2π , and $2\pi n k A_{ij}$ is an integer multiple of 2π it follows that

$$\begin{aligned}
 \sum_{i=1}^n \cos(\theta_i - \sum_j \hat{\beta}_j x_{ij}) &= \sum_{i=1}^n \cos \left(\theta_i - \sum_j \hat{\beta}_j^k (A_{ij} \delta_j) \right) \\
 &= \sum_{i=1}^n \cos(\theta_i - \sum_j \hat{\beta}_j^k x_{ij}),
 \end{aligned}$$

which implies that $\hat{\beta}_j^k$ is also a maximum likelihood estimator of β_j . Since, in practice, covariate values are likely to be recorded to only limited decimal accuracy, this in turn implies that the problem of non-uniqueness of the maximum likelihood estimates can affect almost any data analysis using Gould's model.

Along with the incidence of an infinite number of global maxima, local maxima are also common for the Gould model. Due to these difficulties, the model is seldom used in practice.

3.1.2 Fisher and Lee (1992)

Fisher and Lee proposed a set of models based on a set of conditional distributions suggested by Johnson and Wehrly. In particular their models corresponded to the following three models:

1. Modeling the mean direction of a von Mises random variable as a function of the covariate vector \mathbf{x} .
2. Modeling the concentration of a von Mises random variable as a function of the covariate vector \mathbf{x} .
3. Simultaneously modeling both the mean direction and concentration (referred to as the “mixed model”) of a von Mises random variable as a function of the covariate vector \mathbf{x} .

For model 1, Fisher and Lee assumed that all the concentration parameters were equal and that the i^{th} mean is related to the covariates as

$$\mu_i = \mu + g(\mathbf{x}_i' \boldsymbol{\beta}), \quad (3.2)$$

where $g(\cdot)$, referred to as a “link” function, is assumed to be a monotone mapping of the real line to the interval $(-\pi, \pi]$. Fisher and Lee suggested a parametric family of suitable links as

$$g(x) = 2 \tan^{-1}(\text{sgn}(x)|x|^\lambda), \quad (3.3)$$

where $\lambda = 0$ corresponds to a log transformation. The λ parameter is chosen based on the data, in the spirit of the Box-Cox transformation. One such link function used by Fisher and Lee was

$$g(\cdot) = 2 \tan^{-1}(\cdot).$$

(Note that the Gould model in equation 3.1 corresponds to taking g as the identity function modulo 2π , which does not satisfy the assumption of monotonicity required by Fisher and Lee.)

For a single covariate, the mean direction travels around the cylinder exactly once as the covariate varies from $-\infty$ to ∞ . This avoids the identifiability problem of the maximum likelihood estimates for the Gould model; however this model also has difficulties with multimodality of the likelihood. Even though there seems to be

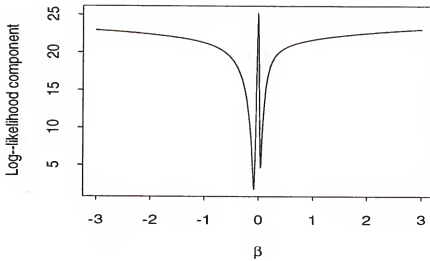


Figure 3.2. Log-likelihood component for the Fisher and Lee model with the periwinkle data.

a single global maximum, there are often a large number of local maxima on the likelihood surface. Fisher and Lee (p. 670-671) recognized that “local maxima can exist quite close to the global maximum so that graphical exploration of the likelihood is advisable anyway.”

Fisher and Lee suggested an iterative Newton-Raphson algorithm, along with graphical exploration of the likelihood, for maximum likelihood parameter estimation. Figure 3.2 shows a plot of the log-likelihood component (the portion of the log-likelihood necessary for estimation of the regression parameters) for the periwinkle data, given in Table 1 of Fisher and Lee and replicated in Table 6.1, which involves only a single predictor. It is easy to see that only for starting values in a very narrow interval would the iterative process converge to the global maximum. Unfortunately, graphical exploration of the likelihood becomes increasingly more difficult as the number of covariates increases. Figure 3.3 shows a plot of the log-likelihood component

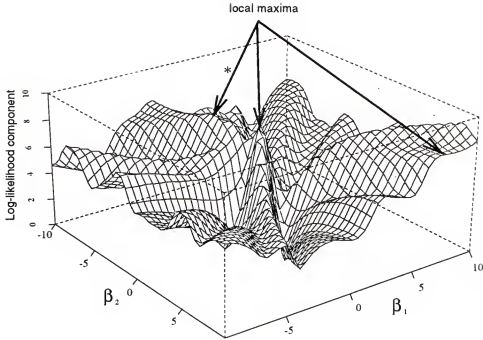


Figure 3.3. Log-likelihood component of the Fisher and Lee model for a simulated data set with $n = 10$, $\kappa = 1.0$, $\beta_1 = .1$ and $\beta_2 = .1$. The global maximum is at *.

for a simulated data set involving two covariates. Graphical assessment alone appears insufficient for finding a starting value that guarantees the algorithm will locate the global maximum.

For model 2, Fisher and Lee assumed that the directional observations were distributed von Mises with constant mean direction μ and the i^{th} concentration parameter was related to the covariates as

$$\kappa_i = h(\mathbf{x}_i' \boldsymbol{\gamma}), \quad (3.4)$$

where $\boldsymbol{\gamma}$ is a vector of regression parameters. The link function h was assumed to be a mapping of the real line to the interval $[0, \infty)$. One such link function used by Fisher and Lee was

$$h(\cdot) = \exp(\cdot).$$

An iterative Newton-Raphson method was also suggested for the estimation of the regression parameters in the concentration model. No complications with estimation

were found for this model. The arbitrariness of the link function is a deficiency with this model; however, Fisher and Lee did give a method for testing the appropriateness of the specified link. (The arbitrariness in the choice of the link function also occurs with the mean model, but Fisher and Lee do not give a test for the appropriateness of the link function.)

For model 3, the mixed model, the directional observations were assumed to come from the von Mises distribution with μ_i represented by equation (3.2) and κ_i represented by equation (3.4). The technique for parameter estimation was to combine the two previous methods into a two-stage fitting process. The new algorithm would iterate between the two previously described algorithms, with only slight modifications to the Newton-Raphson equations. It seems as though similar difficulties would arise in estimation for this model as with the mean model; however, the dimensionality of the problem makes graphical assessment difficult.

Because of the many local maxima for the likelihood function of the Fisher and Lee mean model, two alternative estimation techniques were investigated. Both of these techniques are discussed in the literature in the area of optimizing functions in the presence of local maxima(minima).

The first approach was based on the generalized simulated annealing method of Bobachevsky, Johnson and Stein (1986). The function to be maximized (minimized) was taken as the log-likelihood component, the portion of the log-likelihood necessary for the estimation of the regression parameters. The procedure was not totally successful, in part because of the large number of parameters requiring specification. No one set of parameters was in any way consistent for finding the global maximum with the generalized simulated annealing procedure.

Price's algorithm (Price (1977)) is another technique for maximizing a function in the presence of numerous local maxima. A slight modification of the original algorithm found good success when applied directly to the log-likelihood component,

for situations in which graphical exploration would be difficult. In a small simulation study, the procedure was able to find the global maximum more than eighty percent of the time. However, only when the concentration and/or the sample size was large would the modified algorithm guarantee finding the global maximum. Appendix B describes more fully the problem of the model and a simulation study of the effectiveness of Price's algorithm for parameter estimation in the Fisher and Lee mean model.

3.2 Classification/ANOVA Tests

Throughout this section it will be assumed that the data come from a one-way model, although more complicated designs will be discussed. Following the notation for classification data when the response is linear, it is assumed that the directional observations can be denoted as $(\theta_{11}, \dots, \theta_{1n_1}, \dots, \theta_{p1}, \dots, \theta_{pn_p})$ where p is the number of treatments and n_i is the number of replicates for the i^{th} treatment with $\sum n_i = N$. Data points coming from the same treatment are assumed to have the same mean direction. The concentration parameter is also assumed to be constant for all treatments. The following procedures are methods for testing if the mean directions for the different treatments are equal.

3.2.1 Watson and Williams (1956)

Watson and Williams proposed a classification test that has a similar form to the ANOVA F-test. Watson and Williams assumed that

$$\theta_{ij} \sim \text{vM}(\mu_i, \kappa),$$

and $n_i = n$. For large κ (it is typically assumed that $\kappa \geq 2$), it can be shown that

$$2\kappa(1 - \cos(\theta_{ij} - \mu_i)) \approx \chi_1^2,$$

and hence,

$$2\kappa \left(n - \sum_{j=1}^n \cos(\theta_{ij} - \bar{\theta}_i) \right) = 2\kappa(n - R_i) \approx \chi_{n-1}^2.$$

Proofs of these results are based on the normal approximation to the von Mises distribution discussed in Section 2.2.1. Similar arguments imply that

$$2\kappa \left(N - \sum_{i=1}^p R_i \right) \approx \chi_{N-p}^2, \quad (3.5)$$

and

$$2\kappa \left(\sum_{i=1}^p R_i - R \right) \approx \chi_{p-1}^2. \quad (3.6)$$

Watson and Williams also argued that these statistics are approximately independent for large κ . Recalling from equation (2.4) that the circular variance for treatment i is proportional to $n - R_i$, equations (3.5) and (3.6) represent a partitioning of the total variability in the data into parts corresponding to treatments and residual. Based on these results Watson and Williams suggested an F-test using

$$\frac{(N-p)(\sum_{i=1}^p R_i - R)}{(p-1)(N - \sum_{i=1}^p R_i)} \approx F_{N-p}^{p-1}. \quad (3.7)$$

An ANOVA table can then be formulated in the same manner as proposed for linear responses.

Stephens (1969) suggested a constant multiplier for this statistic that gives a more accurate approximation to the F distribution in simulation studies. The constant multiplier is

$$1 + \frac{3}{8\hat{\kappa}},$$

where $\hat{\kappa} = A^{-1}(\bar{R})$. This modification was said to give a good approximation for $\kappa \geq 1$.

Stephens (1982) and Harrison, Kanji and Gadsen (1986) have described generalizations of this test to more complicated designs. For a two-way model, with interaction, the interaction term corresponds to

$$\sum_{i=1}^p \sum_{j=1}^n R_{ij} - \sum_{i=1}^p R_i - \sum_{j=1}^n R_j + R.$$

This term, however, is not necessarily positive, and cannot be viewed as a partition of the total variability. Similar problems occur with higher order models. Thus the Watson and Williams is unusable for higher order models in the presence of interaction. However, the Watson and Williams test is very useful for the simple one-way ANOVA setting and also the two-way ANOVA setting without interaction and is commonly used in practice with the adjustment made by Stephens.

3.2.2 Harrison, Kanji and Gadsen (1986)

Harrison, Kanji and Gadsen (1986) observed the difficulties with the Watson and Williams test for more complicated designs and suggested an alternative method that was based on the Euclidean distance between resultant vectors. Harrison, Kanji and Gadsen also assumed that

$$\theta_{ij} \sim \text{VM}(\mu_i, \kappa).$$

Their test statistic is very similar in appearance to the Watson and Williams statistic. It takes the form

$$\delta \frac{(N-p)(\sum_{i=1}^p R_i^2/n - R^2/N)}{(p-1)(N - \sum_{i=1}^p R_i^2/n)} \approx F_{N-p}^{p-1}. \quad (3.8)$$

The term δ is a constant multiplier along the lines of Stephens that has the form

$$\delta = 1 - \frac{1}{5\hat{\kappa}} - \frac{1}{10\hat{\kappa}^2}$$

and is used to improve the approximation to the F distribution.

Harrison and Kanji (1988) then generalized this approach to allow for more complicated designs. The general idea, however, remains the same and is based on squared resultant lengths. One complication lies in the interpretation of interaction. The definition of interaction given in Harrison, Kanji and Gadsen and Harrison and Kanji is only valid for equally replicated treatments. When the treatment groups are not equally replicated, the interpretations of interaction do not remain the same. This is not discussed by Harrison, Kanji and Gadsen, and hence the form of the test only seems applicable when the treatments are equally replicated.

Although not fully discussed by Harrison, Kanji and Gadsen, this test is a simultaneous test of both mean directions and concentrations of treatments. Anderson and Wu recognized this fact and showed a complication of the model through the following simple example. Assume that all sample mean directions are identical. The test is comparing the resultant lengths of the treatment groups, all pointing in the same direction. These lengths could be highly variable resulting in a significant F-test although there is no evidence of differing locations. This must be kept in mind when applying the test to a data set, so that incorrect interpretations are not made. Significance of the F-test implies that the distributions of directions for the treatments differ, not just the mean directions.

3.2.3 Anderson and Wu (1994b)

Anderson and Wu described an alternative method of analysis and gave simulation results comparing it to the two previously mentioned tests. Anderson and Wu assumed

$$\theta_{ij} \sim \text{vM}(\mu_i, \kappa);$$

however, they include tests of their model for data generated as wrapped normal. The Anderson and Wu model can be viewed as the Likelihood Ratio Test assuming the parameter κ is fixed and known. Anderson and Wu replaced κ by its maximum

likelihood estimator $\hat{\kappa}$. In particular the test is of the form

$$2\hat{\kappa}\left(\sum_{i=1}^p R_i - R\right) \approx \chi_{p-1}^2. \quad (3.9)$$

The above test is proportional to the numerator of the Watson and Williams test statistic. In essence, the Anderson and Wu and Watson and Williams tests correspond to the directional analogs of the χ^2 and F-tests for differences between normal means with known and unknown variances, respectively. As expected, the Anderson and Wu test has very similar properties to the Watson and Williams test. A generalization for more complicated designs was also suggested by Anderson and Wu that has a similar form to that above. This generalization also has a reasonable interpretation for interaction amongst factors by a simple extension.

One obvious disadvantage to this test is the use of $\hat{\kappa}$. This parameter is known to be biased upwardly in many situations, implying a high Type I error rate. In the simulation studies described by Anderson and Wu, the parameter κ was assumed to be large, $\kappa \geq 2$. In many settings the true κ parameter is less than two and no information concerning the acceptability of this test is discussed. Further simulation studies may be required to give broader acceptance of this test.

Anderson and Wu gave the results of some simulation studies comparing the Type I error rate and power of their test, the Watson and Williams test, and the Harrison, Kanji and Gadsen test. The results of the simulation study show the behavior of the Watson and Williams and Anderson and Wu tests were very similar. The Harrison, Kanji and Gadsen test tended to be too conservative, in terms of Type I error rate, and was not as powerful as the Watson and Williams and Anderson and Wu tests.

3.2.4 Conclusions

All three ANOVA methods are based on comparisons of the variability in a single treatment group to the overall variability in the data. Although not strictly stated

in all of the texts, there is a fundamental assumption concerning equally replicated treatments. This can most easily be seen by the forms of the Watson and Williams and Anderson and Wu tests, where the resultant lengths of each treatment group are not scaled by their respective sample size. These test statistics, when the sample sizes are unequal, will not have the assumed approximate distributions. The Harrison, Kanji and Gadsen test appears to allow for unequally replicated treatments, but the definition that Harrison, Kanji and Gadsen give for interaction is only valid for equally replicated treatments. Hence, it does not seem as though any of these techniques can be usefully applied when the treatments are not equally replicated.

The regression technique of Gould when applied to a classification setting would correspond to the Watson and Williams test, and the approximate tests given by Gould correspond to those of Watson and Williams in the classification setting. There are also difficulties in estimation with the Fisher and Lee regression technique when applied to a classification setting. Due to the discontinuity of points on opposite sides of the circle from the mean, the model can have complications with certain data sets.

CHAPTER 4 A NEW REGRESSION MODEL

4.1 Introduction and Motivation

Directional data are not truly one nor two dimensional, but lie somewhere between these two dimensions. The regression models discussed in Chapter 3 were attempts to model the directional response in one dimension, or to linearize the directional response. Specifically, Gould's model corresponds to a model on the real line that is wrapped onto the unit circle and Fisher and Lee's model attempts to project the unit circle onto the real line after a rotation. A new model is proposed that results from some bivariate assumptions.

Motivation for the new model comes from a simple data collection scenario. Imagine an experimenter measuring the directional travel of birds, after release, in response to some stimuli. The experimenter records the direction of the bird when the bird travels out of site of the experimenter, often referred to as the vanishing angle. The actual distance at which the bird leaves the sight of the experimenter will vary from bird to bird, and depend on things like cloud cover and orientation of sun with respect to the vanishing angle of the bird, in addition to the measured stimuli.

Figure 4.1 shows an example of this idea. The plotted points represent the combination of vanishing angle and distance from observer when the bird disappears from view, in response to the stimuli. The experimenter records only the direction of each point. The distance from observer is something that is usually impossible to measure. However, if it was possible to fit a parametric curve through these points as a function of the covariate (stimuli) this would result in a corresponding model for

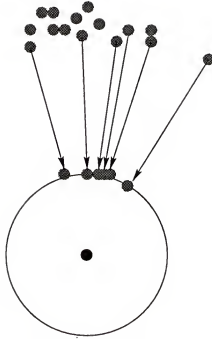


Figure 4.1. Actual data in two-space, where only direction is recorded.

the directional observations alone. Figure 4.2 shows an example of a curve through the data as a function of the covariate (stimuli) and the corresponding model for the directional observations.

4.2 Model Specification

The new model's formulation follows the arguments above by specifying distributional and parametric assumptions as needed. The underlying distribution of the bivariate process is assumed to be bivariate normal. The mean vector of the bivariate normal is modeled through a pair of linear functions of the covariate and the covariance matrix is assumed to be the identity matrix. These assumptions result in the marginal distribution of the direction being offset normal with mean direction represented by the direction of the bivariate normal mean vector and concentration parameter represented by the distance of the bivariate normal mean from the origin.

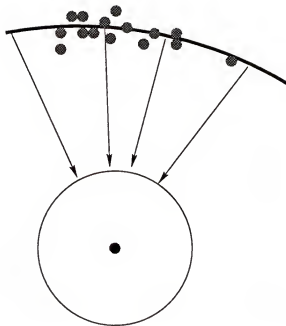


Figure 4.2. Fitting of the mean in two-space, and its associated model of the mean direction.

Symbolically, the new model can be generated as follows. Let us assume

$$\begin{pmatrix} L_i \cos(\theta_i) \\ L_i \sin(\theta_i) \end{pmatrix} = \begin{pmatrix} y_{1i} \\ y_{2i} \end{pmatrix} \sim BN \begin{pmatrix} \mathbf{x}'_i \boldsymbol{\beta}_1 \\ \mathbf{x}'_i \boldsymbol{\beta}_2 \end{pmatrix}, \quad \mathbf{I}. \quad (4.1)$$

It follows from Chapter 2 that

$$\theta_i \sim \text{ON}(\mu_i, \kappa_i), \quad (4.2)$$

where

$$\mu_i = \tan^{-1}(\mathbf{x}'_i \boldsymbol{\beta}_1, \mathbf{x}'_i \boldsymbol{\beta}_2) \quad (4.3)$$

and

$$\kappa_i = [(\mathbf{x}'_i \boldsymbol{\beta}_1)^2 + (\mathbf{x}'_i \boldsymbol{\beta}_2)^2]^{\frac{1}{2}}. \quad (4.4)$$

The function $\tan^{-1}(a, b)$ is defined in equation (2.3).

It is important to realize that the generation described above is not necessary; however, it allows for an interpretation of the form of the model and will motivate

a useful technique for estimation. Equations (4.2)-(4.4) alone constitute the model, and they imply that both the mean direction and the concentration parameter are related to the covariates.

Defining

$$\beta_i = (\beta_{0i}, \dots, \beta_{pi})',$$

and assuming specific forms for the linear functions of the covariates, many different settings can be modeled.

4.2.1 Mean and Concentration Only

By allowing only an intercept in each linear term (the numerator term and the denominator term), the resulting directional model corresponds to

$$\mu_i = \tan^{-1}(\beta_{01}, \beta_{02}), \quad (4.5)$$

$$\kappa_i = \sqrt{\beta_{01}^2 + \beta_{02}^2}. \quad (4.6)$$

This is a model which involves only two parameters, that are one-to-one functions with the mean direction and concentration. Thus, the formulation above allows for “the simplest” model, a constant mean and concentration model. This model will be important for testing the goodness of fit for higher order models where this will be taken as the null-model.

4.2.2 Single Covariate Model

By allowing an intercept and slope in each linear term, the resulting directional model corresponds to

$$\mu_i = \tan^{-1}(\beta_{01} + \beta_{11}x_i, \beta_{02} + \beta_{12}x_i),$$

$$\kappa_i = \sqrt{(\beta_{01} + \beta_{11}x_i)^2 + (\beta_{02} + \beta_{12}x_i)^2}.$$

This model allows for both the mean direction and concentration to be functions of the covariate. For a single covariate traveling from $-\infty$ to ∞ , the mean direction travels exactly half-way around the circle. This model is useful when the mean direction is believed to wrap less than half-way around the circle as the covariate is changing. As the covariate approaches $\pm\infty$, the estimated regression model tends to degenerate distributions, because the concentration parameter tends to ∞ .

4.2.3 Models Wrapping Completely Around the Circle

The model described in Section 4.2.2 would seem to have limited appeal due to the fact that the mean direction can only travel half-way around the circle. This difficulty can be overcome by allowing for quadratic terms in each linear function. The corresponding directional model, given by

$$\begin{aligned}\mu_i &= \tan^{-1}(\beta_{01} + \beta_{11}x_i + \beta_{21}x_i^2, \beta_{02} + \beta_{12}x_i + \beta_{22}x_i^2), \\ \kappa_i &= \sqrt{(\beta_{01} + \beta_{11}x_i + \beta_{21}x_i^2)^2 + (\beta_{02} + \beta_{12}x_i + \beta_{22}x_i^2)^2},\end{aligned}$$

can have a number of different shapes including wrapping fully around the circle. This method also allows for what are termed “returning models.” By this it is meant, mean models whose mean direction goes in one direction and then reverses direction. Figure 4.3 shows some plots of possible models for a single covariate. The directional mean is plotted on the Y -axis, and the covariate is plotted on the X -axis. Due to the circular nature of the response, the top and bottom of this plot are close to one another. The Y -axis corresponds to a circular variate, hence the true plots would actually correspond to curves on the surface of a cylinder with radius one. The X -axis of these curves can be expanded or contracted so that the plotted curves can have a wide range of shapes. The purpose of these curves is to give some idea of the broad settings that the new model can be used for. In many applications, these

higher order curves are not necessary. As the covariate approaches $\pm\infty$, the model tends to a point distribution, since the concentration parameter is tending to ∞ .

4.2.4 Concentration Only

By modeling only a slope parameter and no intercept in each of the linear functions, the corresponding directional model, given by

$$\begin{aligned}\mu_i &= \tan^{-1}(\beta_{11}x_i, \beta_{12}x_i) = \tan^{-1}(\beta_{11}\text{sgn}(x_i), \beta_{12}\text{sgn}(x_i)), \\ \kappa_i &= \sqrt{(\beta_{11}x_i)^2 + (\beta_{12}x_i)^2} = |x_i| \sqrt{\beta_{11}^2 + \beta_{12}^2},\end{aligned}$$

is a model whose mean direction is constant over the positive and negative ranges of the covariate, but on opposite sides of the circle. The concentration parameter is modeled as a constant multiple of the absolute value of the covariate. When the covariate is zero, the model corresponds to the uniform distribution. If the data were strictly positive, or made strictly positive by a location adjustment, the model would correspond to a concentration only model. The mean would be constant throughout the covariate range, and the concentration would be a constant multiple of the covariate. There are many settings in which it is reasonable to assume that when the covariate is zero the directional response is uniform. A simple example of this phenomenon would correspond to modeling wind direction as a function of wind speed. When wind speed is zero, there will be no information about wind direction.

4.2.5 ANOVA Setting

By using an appropriate design matrix, as for the linear response case, the resulting classification model will result in a separate mean direction and concentration being fit for each treatment. Using the same notation as in Section 3.2 the model can be

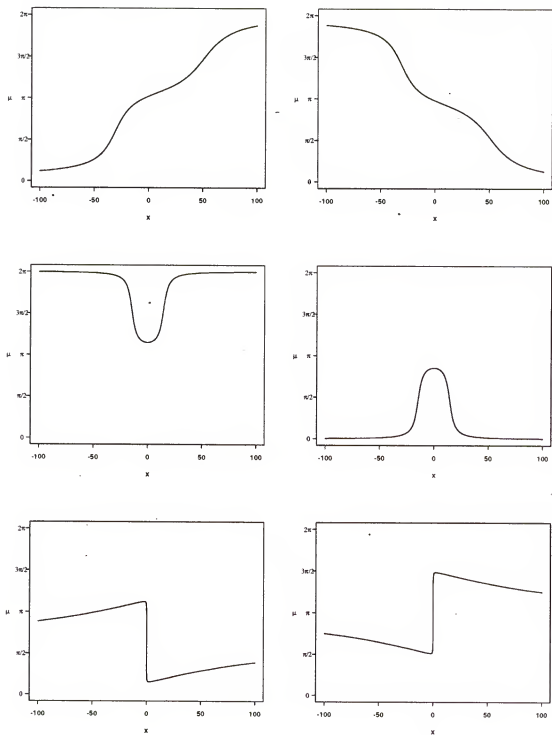


Figure 4.3. Possible curves for the mean direction, by allowing quadratic terms in the model.

written as

$$\begin{aligned}\mu_{ij} &= \tan^{-1}(\beta_{01i}, \beta_{02i}), \\ \kappa_{ij} &= \sqrt{\beta_{01i}^2 + \beta_{02i}^2}.\end{aligned}$$

This model is slightly different from traditional ANOVA settings, since the concentration parameter is not assumed to be constant for all treatment groups.

4.3 Estimation

The derivatives of the log-likelihood for the new model with respect to the parameter vectors β_1 and β_2 take the form

$$-X'X\beta_1 + X'LC \quad (4.7)$$

$$-X'X\beta_2 + X'LS, \quad (4.8)$$

where L is a diagonal matrix with diagonal elements

$$\kappa_i \cos(\theta_i - \mu_i) + \frac{\sqrt{2\pi} \exp\left(\frac{1}{2}\kappa_i^2 \cos^2(\theta_i - \mu_i)\right) \Phi(\kappa_i \cos(\theta_i - \mu_i))}{1 + \sqrt{2\pi}\kappa_i \cos(\theta_i - \mu_i) \exp\left(\frac{1}{2}\kappa_i^2 \cos^2(\theta_i - \mu_i)\right) \Phi(\kappa_i \cos(\theta_i - \mu_i))}. \quad (4.9)$$

In equation (4.9), $\Phi(\cdot)$ is the standard normal CDF, κ_i and μ_i are defined by equations (4.3) and (4.4), and

$$C = \begin{bmatrix} \cos \theta_1 \\ \vdots \\ \cos \theta_n \end{bmatrix}, \quad S = \begin{bmatrix} \sin \theta_1 \\ \vdots \\ \sin \theta_n \end{bmatrix}. \quad (4.10)$$

Unfortunately, but not unexpectedly, there are no closed form solutions for the parameter estimates and an iterative estimation technique is required.

4.3.1 The EM Algorithm

Dempster, Laird and Rubin (1977) unified an approach for parameter estimation discussed by several authors in special cases. This approach, called the EM algorithm, assumes that the observed data can be viewed as having both complete and incomplete specifications (likelihood forms). The observed data are referred to as the incomplete data. Typically the likelihood equations for the observed data are complicated and the estimating equations do not have closed form. The complete data likelihood is usually assumed to be of some rather simple structure with closed form estimating equations.

The EM algorithm has two distinct steps. The expectation (E) step on the p^{th} iteration involves estimating the sufficient statistics for the complete data specification by their expectation, conditional on the incomplete data and the previous iteration's estimated parameters. The maximization (M) step on the p^{th} iteration involves maximizing the complete data likelihood using the most recent estimated values of the sufficient statistics. The algorithm iterates between these steps until convergence.

To formalize this idea, some notation of Dempster, Laird and Rubin will be defined and an example is described which was used in their paper. It is assumed that there exist two sample spaces, \mathcal{Y} and \mathcal{X} , and a many-to-one mapping from \mathcal{X} to \mathcal{Y} . The observed data Y come from \mathcal{Y} . The corresponding X from \mathcal{X} is not fully observed, but must have a mapping to the corresponding Y value. Hence, X lies in a subset of \mathcal{X} , denoted $\mathcal{X}(Y)$. X is referred to as the complete data and Y as the incomplete data. It is assumed that both the likelihoods of X and Y depend on the same set of parameters β . The complete data specification is denoted by $f(X|\beta)$, and the incomplete data specification is denoted by $g(Y|\beta)$. The complete data specification is related to the incomplete data specification through

$$g(Y|\beta) = \int_{\mathcal{X}(Y)} f(X|\beta) dX. \quad (4.11)$$

The EM algorithm itself is defined through the following two steps. The estimation(E) step for the p^{th} iteration corresponds to estimating the sufficient statistics (denoted $t(X)$) of the complete data specification by their expectation, conditional upon the incomplete data, Y , and the previous iterations parameter estimates. Denoting this as $t^{(p)}$, gives

$$t^{(p)} = \mathbf{E}(t(X)|Y, \hat{\beta}^{(p-1)}).$$

The maximization(M) step for the p^{th} iteration then corresponds to maximizing the complete data likelihood, X , using the estimated sufficient statistics. The estimate of the parameter at this iteration is denoted as $\hat{\beta}^{(p)}$. The form of these equations are not always simple; however, the EM algorithm is best utilized in situations when the complete data likelihood has simple maximum likelihood estimates. In particular, when the complete data specification is of the regular exponential family, it follows that $\hat{\beta}^{(p)}$ is the solution to

$$\mathbf{E}(t(X)|\beta) = t^{(p)}.$$

This corresponds to the maximum likelihood estimating equations for the regular exponential family. The algorithm iterates between these two steps until convergence.

The simplest example of an application of the EM algorithm is discussed in Dempster, Laird and Rubin's examples section. In the analysis of variance setting for normally distributed responses with known variance, the maximum likelihood estimates of parameters have simple forms, corresponding to differences between treatment means, as long as the design is balanced. When the design matrix is unbalanced, maximum likelihood estimation requires the inversion of a large matrix. Instead of inverting the matrix, a balanced design can also be achieved by replacing the missing responses by their expectations conditional upon the current parameter estimates (E-step). The simple form of the estimating equations could then be used to estimate the parameters (M-step). Iteration between these two steps until the parameter estimates

converge yields the maximum likelihood estimates of the parameters. In fact, it is possible to add rows to the design matrix that were not originally present to allow for simpler estimating equations. The EM algorithm is useful and most often applied in this setting with missing data; however this is not the only situation in which the EM can be applied.

4.3.2 Application to the New Model

The formulation of the new model provides natural complete and incomplete data specifications. The incomplete data are the observed directions and the complete data (unobserved) are the combination of direction and distance from the origin. The complete data specification corresponds to the bivariate normal and the incomplete data specification is the offset normal described through equations (4.1)-(4.4). The imagined bivariate vectors, defined as \mathbf{Y}_1 and \mathbf{Y}_2 , can be written as

$$\mathbf{Y}_1 = \begin{bmatrix} L_1 \cos \theta_1 \\ \vdots \\ L_n \cos \theta_n \end{bmatrix}, \quad \mathbf{Y}_2 = \begin{bmatrix} L_1 \sin \theta_1 \\ \vdots \\ L_n \sin \theta_n \end{bmatrix}.$$

The sufficient statistics for the complete data specification are $\mathbf{X}'\mathbf{Y}_1$ and $\mathbf{X}'\mathbf{Y}_2$. (The variance-covariance matrix of both \mathbf{Y}_1 and \mathbf{Y}_2 are assumed known.) Maximizing the likelihood of the complete data specification corresponds to the familiar ordinary least squares equations

$$\hat{\beta}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_1$$

$$\hat{\beta}_2 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_2$$

For the new model the two steps of the EM algorithm are carried out as follows:

1. The E step for the p^{th} iteration is executed by estimating $\mathbf{X}'_1\mathbf{Y}_1$ and $\mathbf{X}'_2\mathbf{Y}_2$ by

$$\mathbf{X}'_1\hat{\mathbf{Y}}_1^p = \mathbf{X}'_1\mathbf{E}(\mathbf{Y}_1|\Theta, \hat{\beta}_1^{p-1}, \hat{\beta}_2^{p-1}),$$

and

$$\mathbf{X}_2' \widehat{\mathbf{Y}}_2^p = \mathbf{X}_2' \mathbf{E}(\mathbf{Y}_2 | \Theta, \widehat{\beta}_1^{p-1}, \widehat{\beta}_2^{p-1}),$$

where

$$\mathbf{E}(\mathbf{Y}_1 | \Theta, \widehat{\beta}_1^{p-1}, \widehat{\beta}_2^{p-1}) = \begin{cases} \cos(\theta_1) \mathbf{E}(L_1 | \Theta, \widehat{\beta}_1^{p-1}, \widehat{\beta}_2^{p-1}) \\ \cos(\theta_2) \mathbf{E}(L_2 | \Theta, \widehat{\beta}_1^{p-1}, \widehat{\beta}_2^{p-1}) \\ \dots \\ \cos(\theta_n) \mathbf{E}(L_n | \Theta, \widehat{\beta}_1^{p-1}, \widehat{\beta}_2^{p-1}) \end{cases}$$

The representation for $\mathbf{E}(\mathbf{Y}_2 | \Theta, \widehat{\beta}_1^{p-1}, \widehat{\beta}_2^{p-1})$ is identical to that above with all cosine terms replaced by their corresponding sine terms. The conditional expectation of L_i , $\mathbf{E}(L_i | \Theta, \widehat{\beta}_1^{p-1}, \widehat{\beta}_2^{p-1})$ can be expressed (see Appendix C) in closed form as

$$\widehat{\kappa}_i \cos(\theta_i - \widehat{\mu}_i) + \frac{\sqrt{2\pi} \exp\left(\frac{1}{2} \widehat{\kappa}_i^2 \cos^2(\theta_i - \widehat{\mu}_i)\right) \Phi(\widehat{\kappa}_i \cos(\theta_i - \widehat{\mu}_i))}{1 + \sqrt{2\pi} \widehat{\kappa}_i \cos(\theta_i - \widehat{\mu}_i) \exp\left(\frac{1}{2} \widehat{\kappa}_i^2 \cos^2(\theta_i - \widehat{\mu}_i)\right) \Phi(\widehat{\kappa}_i \cos(\theta_i - \widehat{\mu}_i))}, \quad (4.12)$$

where $\widehat{\mu}_i$ and $\widehat{\kappa}_i$ are the estimates based on $\widehat{\beta}_1^{p-1}$ and $\widehat{\beta}_2^{p-1}$. A convenient notation is to write

$$\mathbf{X}_1' \widehat{\mathbf{Y}}_1^p = \mathbf{X}_1' \widehat{\mathbf{L}}^p \mathbf{C}, \quad (4.13)$$

$$\mathbf{X}_2' \widehat{\mathbf{Y}}_2^p = \mathbf{X}_2' \widehat{\mathbf{L}}^p \mathbf{S}, \quad (4.14)$$

where $\widehat{\mathbf{L}}^p$ is a diagonal matrix with elements

$$\mathbf{E}(L_i | \Theta, \widehat{\beta}_1^{p-1}, \widehat{\beta}_2^{p-1}),$$

\mathbf{C} is the vector of cosines defined previously, and \mathbf{S} is the corresponding vector of sines defined previously.

2. The M step for the p^{th} iteration amounts to maximizing the full likelihood of the complete data specification using the estimated sufficient statistics in equations

(4.13) and (4.14). This leads to the familiar equations

$$\hat{\beta}_1^p = (X_1' X_1)^{-1} X_1' \hat{Y}_1^p = (X_1' X_1)^{-1} X_1' \hat{L}^p C \quad (4.15)$$

$$\hat{\beta}_2^p = (X_2' X_2)^{-1} X_2' \hat{Y}_2^p = (X_2' X_2)^{-1} X_2' \hat{L}^p S. \quad (4.16)$$

Equations (4.15) and (4.16) correspond to the OLS equations with the response vector replaced by $\hat{L}^p C$ and $\hat{L}^p S$, respectively. By realizing this fact it is possible to use standard statistical packages during each iteration of the maximization step. Equations (4.15) and (4.16) have the same form as the scoring equations generated by equations (4.7) and (4.8) with the L matrix replaced by an estimate. The algorithm requires us to choose a starting value for \hat{L} , but the choice is arbitrary. In practice $\hat{L}^1 = I$ has been used with good results.

4.4 Inference

4.4.1 Global Maximality of Estimates

Much discussion concerning the EM algorithm has centered around the presence of local maxima in addition to the global maximum on the likelihood surface. (See Wu (1983).) Dempster, Laird and Rubin give a proof that the likelihood function is a nondecreasing function from iteration to iteration. Because of this fact, the EM algorithm guarantees finding the global maximum on the likelihood surface if the negative of the Hessian matrix, defined as the matrix of second partials of the log-likelihood, is non-negative definite. The Hessian matrix for the combined vector $(\beta_1' : \beta_2')$ of the new model can be written in the form

$$\begin{bmatrix} X' & 0 \\ 0 & X' \end{bmatrix} V \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix}, \quad (4.17)$$

where

$$V = \begin{bmatrix} I - WC^2 & -WCS \\ -WCS & I - WS^2 \end{bmatrix}. \quad (4.18)$$

In equation (4.18), CS is a diagonal matrix with diagonal elements $\cos(\theta_i)\sin(\theta_i)$, C^2 and S^2 are diagonal matrices with diagonal elements $\cos^2(\theta_i)$ and $\sin^2(\theta_i)$ respectively, and W is a diagonal matrix whose elements are

$$w_{ii} = \left(2 - \frac{\sqrt{2\pi}Q_i\Phi(Q_i)\exp\left(\frac{1}{2}Q_i^2\right)}{1 + \sqrt{2\pi}Q_i\Phi(Q_i)\exp\left(\frac{1}{2}Q_i^2\right)} - \frac{2\pi\Phi^2(Q_i^2)\exp(Q_i^2)}{\left\{1 + \sqrt{2\pi}Q_i\Phi(Q_i)\exp\left(\frac{1}{2}Q_i^2\right)\right\}^2} \right), \quad (4.19)$$

where $Q_i = \kappa_i \cos(\theta_i - \mu_i)$. Appendix D gives a proof showing the negative of the Hessian matrix is non-negative definite for our model, implying the global maximality of our parameter estimates.

4.4.2 Wald-Type Tests

Wald (1949) first gave arguments implying the consistency and asymptotic normality of maximum likelihood parameter estimates. Wald's arguments assume the observations are independent and identically distributed. Under similar regularity conditions, it can be shown that the maximum likelihood estimators $\hat{\beta}_1$, and $\hat{\beta}_2$ are consistent and asymptotically normal in many regression settings. Specifically, by denoting $\beta' = (\beta'_1, \beta'_2)$ and

$$Z = \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix},$$

standard asymptotic theory implies

$$\hat{\beta} \approx \text{MVN}(\beta, I(\hat{\beta})^{-1}), \quad (4.20)$$

where $I(\hat{\beta})$ is the observed information matrix, defined as the negative of the Hessian matrix evaluated at $\hat{\beta}$, i.e.

$$I(\hat{\beta}) = Z' \hat{V} Z.$$

For the model with only an intercept in each linear term, the mean and concentration only model, the observed information matrix can be written as

$$I(\hat{\beta}) = Z' \hat{V} Z = \begin{bmatrix} n - \sum_{i=1}^n \hat{w}_{ii} \cos^2 \theta_i & -\sum_{i=1}^n \hat{w}_{ii} \cos \theta_i \sin \theta_i \\ -\sum_{i=1}^n \hat{w}_{ii} \cos \theta_i \sin \theta_i & n - \sum_{i=1}^n \hat{w}_{ii} \sin^2 \theta_i \end{bmatrix}.$$

The observed information matrix is used for two reasons. First, the expected information matrix requires numerical integration. Second, the observed information matrix also is “slightly more accurate” (Fisher (p. 156) (1973)), in estimating the variance covariance matrix of $\hat{\beta}$. (See Efron and Hinkley (1978) for further discussion and recommendation of the use of the observed information matrix over the expected information matrix.)

Remark:

In all of the above discussions, the model for both linear equations is always taken to be the same. The rationale behind this concerns the problem of rotation. Specifically, assume that all of the data are rotated by α , i.e.

$$\theta_i^\alpha = (\theta_i - \alpha) \bmod 2\pi.$$

The estimating equations would then take the form

$$\hat{\beta}_1^\alpha = (X'X)^{-1} X' \hat{L} C^\alpha, \quad (4.21)$$

$$\hat{\beta}_2^\alpha = (X'X)^{-1} X' \hat{L} S^\alpha. \quad (4.22)$$

Using simple trigonometric identities, equations (4.21) and (4.22) can be written as

$$\hat{\beta}_1^\alpha = (X'X)^{-1} X' \hat{L} [C(\cos \alpha) + S(\sin \alpha)],$$

$$\hat{\beta}_2^\alpha = (X'X)^{-1} X' \hat{L} [S(\cos \alpha) - C(\sin \alpha)],$$

which imply

$$\begin{aligned}\hat{\beta}_1^\alpha &= \cos \alpha \hat{\beta}_1 + \sin \alpha \hat{\beta}_2, \\ \hat{\beta}_2^\alpha &= \cos \alpha \hat{\beta}_2 - \sin \alpha \hat{\beta}_1.\end{aligned}$$

Thus, if a higher order term is not included in one of the linear terms, by a simple rotation of the data, these higher order terms could then be included. Hence, it makes more sense to include the same terms in both linear equations, even if some terms are negligible. When testing significance of a factor, it is important to simultaneously test both the terms in the numerator and the denominator corresponding to that factor. Significance of these tests imply that the factor is significantly affecting the fit of the model. This effect could involve the mean fit and/or the concentration fit, hence the tests are for differing distributions.

4.4.3 Likelihood-Ratio Tests

Exact and asymptotic tests based on the likelihood ratio statistic are common techniques used for hypothesis testing. Defining $L(\beta)$ as the likelihood of the response vector $\theta_1, \dots, \theta_n$, tests for the hypotheses

$$H_0 : \beta \in \beta_0 \quad \text{vs.} \quad H_A : \beta \in \beta_A,$$

where β_0 is a subset of β_A , are based on

$$\Lambda = \frac{\sup_{\beta \in \beta_0} L(\beta)}{\sup_{\beta \in \beta_A} L(\beta)}.$$

Specifically, the null hypothesis is rejected for Λ small. Under similar (identical) regularity conditions necessary for the Wald-type tests, it can be shown that asymptotically

$$-2 \log \Lambda \approx \chi_{n_a - n_0}^2,$$

where n_a is the number of free parameters in H_A and n_0 is the number of free parameters in H_0 .

4.4.4 Approximate Confidence Intervals

The techniques in Sections 4.4.2 and 4.4.3 allow for assessing the significance of factors in our model. It is also of interest to use the parameter estimates for estimation and prediction of the response. By substituting in the parameter estimates, simple estimates of the mean fit can be found. Methods of assessing the goodness of fit of this mean fit are also required. The delta method allows us to estimate standard errors of the mean fit. (See Agresti(1990, p. 419-424) for a good review.) In general the delta method implies the asymptotic normality of functions of asymptotically normal statistics.

Defining

$$\mu_x = \tan^{-1}(\mathbf{x}'\beta_1, \mathbf{x}'\beta_2),$$

and assuming that

$$\hat{\beta} \approx \text{MVN}(\beta, (\mathbf{Z}'\widehat{\mathbf{V}}\mathbf{Z})^{-1}),$$

and

$$\psi_{xk} = \left. \frac{\partial \mu_x}{\partial \beta_k} \right|_{\beta} \neq 0,$$

it follows from the delta method that

$$\hat{\mu}_x \approx N(\mu_x, \hat{\Psi}'_x(\mathbf{Z}'\widehat{\mathbf{V}}\mathbf{Z})^{-1}\hat{\Psi}_x),$$

where $\Psi'_x = (\psi_{x1}, \dots, \psi_{x2p})$. This allows us to construct approximate pointwise confidence intervals that take the form

$$\tan^{-1}(\mathbf{x}'\hat{\beta}_1, \mathbf{x}'\hat{\beta}_2) \pm z\sqrt{\hat{\Psi}_x(\mathbf{Z}'\widehat{\mathbf{V}}\mathbf{Z})^{-1}\hat{\Psi}_x},$$

for the mean fit when the sample size is large.

4.4.5 Approximate Prediction Bounds

One approach for constructing prediction bounds can be accomplished by employing the standard techniques for linear variates. Following the formulation in linear responses, define $\psi = \theta_o - \hat{\mu}_o$, where θ_o is some future observation and $\hat{\mu}_o$ is the associated estimate of θ_o . The variance of ψ is

$$\mathbf{Var}(\theta_o - \hat{\mu}_o) = \mathbf{Var}(\theta_o) + \mathbf{Var}(\hat{\mu}_o),$$

since the future observation θ_o is independent of $\hat{\mu}_o$. For inference about future observations, the variance of ψ is the correct variance to use. Hence, the variance of a future observation, is the combination of the variance of the estimator plus the variance associated with a single observation. From Section 4.4.4

$$\widehat{\mathbf{Var}}(\hat{\mu}_o) = \hat{\Psi}'_o(\mathbf{Z}'\widehat{\mathbf{V}}\mathbf{Z})^{-1}\hat{\Psi}_o.$$

However, an estimate of $\mathbf{Var}(\theta_o)$ is still required

Mardia gives an alternative representation for any circular density function as

$$f(\theta) = \frac{1}{2\pi} \left[1 + 2 \sum_{p=1}^{\infty} \{ \alpha_p \cos(p\theta) + \beta_p \sin(p\theta) \} \right], \quad (4.23)$$

where α_p and β_p are the trigonometric moments defined in Appendix A. Expanding the cosine and sine terms in equation (4.23) gives

$$f(\theta) = \frac{1}{2\pi} \left[1 + 2 \sum_{p=1}^{\infty} \alpha_p \left\{ \sum_{j=0}^{\infty} \frac{(-1)^j (p\theta)^{2j}}{(2j)!} \right\} + 2 \sum_{p=1}^{\infty} \beta_p \left\{ \sum_{j=0}^{\infty} \frac{(-1)^j (p\theta)^{2j+1}}{(2j+1)!} \right\} \right]. \quad (4.24)$$

This representation of a density permits finding the first and second moments of θ and hence $\mathbf{Var}(\theta_o)$. Term-by-term integration (taking $\theta \in [-\pi, \pi)$) of equation (4.24) implies that

$$\mathbf{E}(\theta^2) = \frac{\pi^2}{3} + 2\pi^2 \sum_{p=1}^{\infty} \alpha_p \left[\sum_{j=0}^{\infty} \frac{(-1)^j (p\pi)^{2j}}{(2j)!(2j+3)} \right],$$

and

$$\mathbf{E}(\theta) = 2\pi \sum_{p=1}^{\infty} \beta_p \left[\sum_{j=0}^{\infty} \frac{(-1)^j (p\pi)^{2j+1}}{(2j+1)!(2j+3)} \right],$$

which imply a closed form for $\mathbf{Var}(\theta)$.

The future observation is assumed to come from the offset normal distribution, whose trigonometric moments are defined in Appendix A. By inserting estimates of α_p and β_p for the particular \mathbf{x}_o of the future observation, found by maximum likelihood estimation, an estimate of the variance of a future observation can be found. Approximate prediction bands for θ_o can then be calculated as

$$\tan^{-1}(\mathbf{x}'_o \hat{\beta}_1, \mathbf{x}'_o \hat{\beta}_2) \pm z \sqrt{\widehat{\mathbf{Var}}(\theta_o) + \hat{\Psi}_x (\mathbf{Z}' \widehat{\mathbf{V}} \mathbf{Z})^{-1} \hat{\Psi}_x}. \quad (4.25)$$

Unfortunately, this prediction band may not be very useful in practice, because it is based on the assumption that the prediction distribution of a directional observation is normal or approximately normal. If this were true, standard multiple linear regression models could be used for the estimation step, and a much simpler model could be found.

Alternatively, a different set of prediction intervals can be constructed. The direction of a future observation, denoted as θ_o , has the form

$$\theta_o = \tan^{-1}(y_{1o}, y_{2o}),$$

where y_{1o} and y_{2o} can be thought of as future observations from the bivariate normal distribution. An estimate of θ_o is $\tan^{-1}(\mathbf{x}_o \hat{\beta}_1, \mathbf{x}_o \hat{\beta}_2) = \tan^{-1}(\hat{y}_{1o}, \hat{y}_{2o})$. Since the vector of $(\hat{\beta}'_1, \hat{\beta}'_2)$ is asymptotically normal, it follows that

$$\begin{pmatrix} \hat{y}_{1o} \\ \hat{y}_{2o} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_o \hat{\beta}_1 \\ \mathbf{x}_o \hat{\beta}_2 \end{pmatrix} \approx BN \left(\begin{pmatrix} \mathbf{x}_o \beta_1 \\ \mathbf{x}_o \beta_2 \end{pmatrix}, \begin{bmatrix} \mathbf{x}'_o \\ \mathbf{x}'_o \end{bmatrix} (\mathbf{Z}' \widehat{\mathbf{V}} \mathbf{Z})^{-1} [\mathbf{x}_o : \mathbf{x}_o] \right),$$

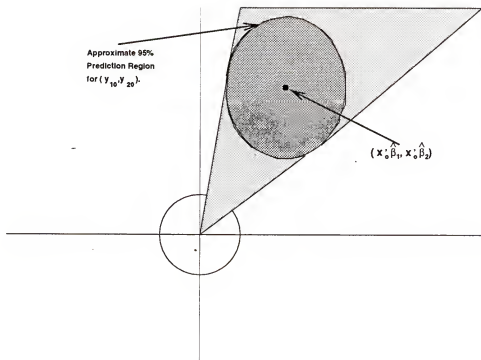


Figure 4.4. Implication of projecting a confidence region in two-space onto the unit circle.

by an application of the delta method. The arguments of prediction variance for the simple regression model imply

$$\begin{pmatrix} y_{10} \\ y_{20} \end{pmatrix} \approx BN \left(\begin{matrix} x'_0 \hat{\beta}_1 \\ x'_0 \hat{\beta}_2 \end{matrix}, \begin{bmatrix} x'_0 \\ x'_0 \end{bmatrix} (Z' \hat{V} Z)^{-1} [x_o : x_o] + I \right). \quad (4.26)$$

The specification in equation (4.26) implies a specific distribution for the predicted direction. Intuitively, the simplest thing to do now would be to construct an approximate prediction interval for the bivariate observation, and then map this onto the unit circle, giving an approximate interval for the mean direction. Unfortunately, this would not be appropriate, as Figure 4.4 implies. The prediction region for a future directional observation would contain all points within the 95% prediction region, denoted by the dark shaded region, as well as all other points within the lightly shaded

region. This prediction interval for the direction will be much wider than necessary, because of including all of these additional points.

Monte-Carlo methods can be incorporated to get more reasonable prediction bands. The general idea is that pseudo-random observations can be generated from the bivariate normal distribution in equation 4.26. Their associated directions can be recorded, and prediction bounds can be based on the distributional properties of these pseudo-random observations. Two different techniques are suggested for estimating the prediction bands. One guarantees symmetric confidence bands, the other does not. It is assumed that N observations ($N > 1000$) are generated for the appropriate value of \mathbf{x}_o . The following methods give approximate $(1 - \alpha)100\%$ prediction intervals for the direction of an observation at \mathbf{x}_o .

The first method can be defined as follows:

- Identify $\hat{\mu}_o = \tan^{-1}(\mathbf{x}'_o \hat{\beta}_1, \mathbf{x}'_o \hat{\beta}_2)$ on the circle and the corresponding value of $(\hat{\mu}_o - \pi) \bmod 2\pi$.
- The parameter γ is then chosen such that

$$\# \text{ observations in the arc defined by } (\hat{\mu}_o - \pi \pm \gamma) = \alpha(N).$$

- An approximate $(1 - \alpha)100\%$ prediction interval can then be calculated as the arc

$$[\hat{\mu}_o \pm (\pi - \gamma)] \bmod 2\pi,$$

which contains $\hat{\mu}_o$.

The second method has the same first step as that above. Instead of choosing endpoints equidistant from $\hat{\mu}_o - \pi$, clockwise and counter clockwise, pick $\frac{\alpha(N)}{2}$ observations in each direction separately that are closest to $\hat{\mu}_o - \pi$. The corresponding arc defined by these endpoints which contains $\hat{\mu}_o$ is then defined to be an approximate $(1 - \alpha)100\%$ prediction interval.

The techniques described above are computationally intensive. A simpler method for obtaining prediction intervals can be used when the estimated concentration for \mathbf{x}_o is large. It has been established that for large κ the von Mises density can be well approximated by the normal density function by taking $1/\kappa$ as the variance (Section 2.2.1). It has also been established that the offset normal distribution and von Mises distribution have very similar shapes for reasonably large κ (Section 2.2.4). This implies that for reasonably large κ values the offset normal observations can be approximated by the normal distribution. Thus an approximate estimate of the variance of θ_o when κ_{x_o} is relatively large is

$$\widehat{\text{Var}}(\theta_o) = \frac{1}{A^{-1}[S(\widehat{\kappa}_{x_o})]},$$

where $A(\cdot)$ is the ratio of Bessel functions defined in Section 2.2.1 and $S(\cdot)$ is as defined in Section 2.2.2. Hence,

$$\tan^{-1}(\mathbf{x}'_o \widehat{\boldsymbol{\beta}}_1, \mathbf{x}'_o \widehat{\boldsymbol{\beta}}_2) \pm z \sqrt{\frac{1}{A^{-1}[S(\widehat{\kappa}_{x_o})]} + \widehat{\Psi}_x(\mathbf{Z}' \widehat{\mathbf{V}} \mathbf{Z})^{-1} \widehat{\Psi}_x}$$

gives an alternative approximate 95% prediction band when the concentration can be assumed large.

4.4.6 Circular R^2 Measures

The terms R^2 and adjusted R^2 for scalar responses correspond to measures of goodness of fit. These quantities measure the change in mean fit for adding or deleting parameters in a regression model. The measures allow for interpretations concerning the improvement of the model. It seems natural to introduce similar measures for directional responses which will be referred to as circular R^2 s.

For scalar responses the terms R^2 and adjusted R^2 can be written as

$$R^2 = 1 - \frac{SSE[F]}{SSE[M]} \quad (4.27)$$

and

$$R_A^2 = 1 - \frac{MSE[F]}{MSE[M]}, \quad (4.28)$$

where $SSE[F]$ ($MSE[F]$) represents the sum of squares (mean square) error for the regression model, and $SSE[M]$ ($MSE[M]$) represents the sum of squares (mean squared) error for the mean only model. Comparable measures can then be assembled using the terms discussed in Chapter 2. Specifically, measures will be constructed using S_0 , the circular variance defined in equation (2.4). Substituting the appropriate circular variance in place of the sums of squares gives us

$$R_{c1}^2 = 1 - \frac{n - \sum_{i=1}^n \cos(\theta_i - \hat{\mu})}{n - \sum_{i=1}^n \cos(\theta_i - \hat{\mu})}, \quad (4.29)$$

where $\hat{\mu}$ is the estimated mean direction from our simplest model, the mean and concentration only model, and $\hat{\mu}_i$ is the i^{th} estimated mean for the proposed regression model. A similar argument leads to an adjusted R^2 by including the degrees of freedom in the model, of the form

$$R_{c3}^2 = 1 - \frac{(n-2)(n - \sum_{i=1}^n \cos(\theta_i - \hat{\mu}_i))}{(n-2p)(n - \sum_{i=1}^n \cos(\theta_i - \hat{\mu}))}. \quad (4.30)$$

A separate pair of circular R^2 statistics can be constructed based on s_0^2 , the squared circular standard deviation defined in equation (2.5). The sums of squares for the scalar R^2 terms are replaced by the squared circular standard deviation to imply

$$R_{c2}^2 = 1 - \frac{\log(\sum_{i=1}^n \cos(\theta_i - \hat{\mu}_i)) - \log n}{\log(\sum_{i=1}^n \cos(\theta_i - \hat{\mu})) - \log n}. \quad (4.31)$$

The final version, which corresponds to an adjusted R^2 takes the form

$$R_{c4}^2 = 1 - \frac{(n-2)[\log(\sum_{i=1}^n \cos(\theta_i - \hat{\mu}_i)) - \log n]}{(n-2p)[\log(\sum_{i=1}^n \cos(\theta_i - \hat{\mu})) - \log n]}. \quad (4.32)$$

It is important to realize that all four of these terms do not necessarily increase as terms are added into the new model, since the parameters are simultaneously affecting both the mean direction and concentration. These functions measure the change in the fit of the mean and do not give any information concerning the improvement in fitting concentration. It is possible and likely that certain terms may significantly benefit the model, but decrease some or all of the circular R^2 measures. By realizing this, it may be possible to separate certain factors in terms of how they are improving the fit of the model. These measures depend only the cosine of the difference between the observed direction and the predicted direction, and are not based on any distributional assumptions related to the observed directions. However, if the data were distributed as von Mises with constant (large) concentration, arguments made by Watson and Williams imply the that R_{c1}^2 and R_{c3}^2 have approximately the same relation to the F-distribution, as the R^2 and adjusted R^2 terms for normally distributed responses.

These measures also will allow for comparisons of different parametric models that are not necessarily nested. This permits for reasonable and useful tests and comparisons between the models proposed by Fisher and Lee, and the new model. The distributional behavior of these statistics will be discussed further in Chapter 5 through simulation studies.

4.4.7 Adjustments for a Mean Model

Most useful models relating to classification or ANOVA models can be thought of as means models. The estimated parameters are based on differences between treatment means. The techniques proposed by Watson and Williams and Anderson and Wu are models of this form. The new model can not be labeled as such, since the maximum likelihood estimate of a treatment mean direction does not correspond

to the treatment's directional means. The length of each observed direction is estimated(imputed) conditioned on the observed direction. Separate lengths are given for each direction within the same treatment group. The resulting bivariate average does not have the same direction as the directional average of the directions alone. (This discussion truly corresponds to the difference between the method of moments and maximum likelihood estimates of the mean direction.) If the resultant length did not depend on the observed direction then the resulting model would be a means model.

By a slight modification in our estimation technique, parameter estimates can be formulated that correspond to the the true directional means. In the estimation step of the EM algorithm, replace equation 4.12 by

$$\mathbf{E}(L_i|\hat{\beta}_1^{p-1}, \hat{\beta}_2^{p-1}) = \sqrt{2} \exp\left(-\frac{1}{2}\hat{\kappa}_i^2\right) \sum_{j=1}^{\infty} \left(\frac{\hat{\kappa}_i^2}{2}\right)^j \frac{\Gamma(j+1.5)}{\Gamma(j+1)}. \quad (4.33)$$

Equation (4.33) differs from equation (4.12) since the direction of the observation is not conditioned on in equation (4.33). Further investigation of this slightly modified estimation technique should be done. It is not clear if this modification will change the asymptotic properties of the parameter estimates.

CHAPTER 5 SIMULATION STUDIES

5.1 Introduction

This chapter is an attempt to test and back up, with simulation studies, some of the theoretical results discussed throughout the text. Specifically, information is sought on three different ideas from the text: distributional properties of the parameter estimates, assessment of the error rates for the Wald and likelihood ratio tests, distributional behavior of the circular R^2 measures. Due to the complexity of even simple models (the simplest regression model that includes a covariate has at least four parameters), only this simple model will be compared to the null model consisting of a constant concentration and constant mean direction.

5.2 Parameter Estimates

Standard statistical techniques imply the asymptotic normality of the parameter estimates under certain regularity conditions. The appropriateness of this approximation generally depends on the combination of the assumed model and the sample size. In order to examine the adequacy of these approximations, a small simulation study was run.

Offset normal random variables were generated as the directions of independent normal random variables generated using the method of Marsaglia and Tsang. The covariate values were generated as Uniform random variables on the interval $(-10, 10)$, and held constant for all parameter choices. One hundred data sets were generated

for each combination of the following parameters:

$$\begin{aligned}
 n &: 20, 50, 100, \\
 \beta_{01} &: .1, 1, \\
 \beta_{02} &: .1, 1, \\
 \beta_{11} &: 0, .3, 1, 3, \\
 \beta_{12} &: 0, .3, 1, 3,
 \end{aligned}$$

where n corresponds to the sample size and β_{0i} and β_{1i} correspond to the intercepts and slopes of the numerator ($i = 2$) and denominator ($i = 1$) term in our model. Because of symmetry conditions discussed in Section 4.4.2, only distributional properties of β_{02} and β_{12} will be discussed. Due to the large number of parameter combinations, an attempt was made to summarize useful information about the parameter estimates in a relatively small amount of space. Figures 5.1-5.3 show sets of side-by-side boxplots of the estimated values of $\hat{\beta}_{12}$ for different settings of the other parameters. Within each individual plot (sixteen plots on each page), the parameter estimates all have the same underlying value of β_{12} , represented by the dashed-horizontal line, and compare the estimates for different values of β_{11} . Comparisons within each individual plot correspond to comparisons of the estimation for β_{12} for differing values of β_{11} .

As can be seen in the plots, there seems to be little difference between the estimates within each separate plot for differing values of β_{11} . This implies little or no effect on the estimation of the slope parameter in the denominator (numerator) term for different values of the slope parameter in the numerator (denominator) term. Comparisons of columns of plots correspond to comparisons of different intercepts for the underlying numerator and denominator parameters. Once again, there seems to be little difference between columns, implying that the behavior of the slope estimates is not affected by differing values of the intercepts, whether in the same term or in

the other term. Comparisons of the different rows correspond to comparisons of the parameter estimates when the value of the slope parameter is changing. In general the average of the parameter estimates seems to be larger than the true value of the parameter and the difference increases with the size of the true parameter. This implies an upward bias in the estimation of the slope parameters. This does not occur when the true value of the slope is zero ($\beta_{12} = 0$). This bias can be observed for all sample sizes, although for samples of size one hundred(100) the bias appears very small.

In general the distributions of the parameters are also upwardly skewed. This can most easily be seen for samples of size 20. In this setting, the boxplots do not show any parameter estimates as being outliers below the true value; however, many parameter estimates are outliers above the true parameter value. As expected, the amount of skewness seems to decrease as the sample size increases. This skewness was not found when the true slope value was zero ($\beta_{12} = 0$).

The skewness and bias in the estimation of the slope parameters is not totally surprising. The small simulation study discussed in Section 2.2.5 demonstrated the upward bias in the estimation of the concentration parameter for the offset normal distribution. The parameters in the new regression model are simultaneously modeling the mean direction and concentration, so it is not at all surprising that these regression parameters are biased upwardly and skewed upwardly.

Figures 5.4-5.6 show side-by-side boxplots of the estimated values of $\hat{\beta}_{02}$ for different settings of the underlying parameters. Within each individual plot (sixteen plots on each page), the parameter estimates all have the same underlying value of β_{02} , represented by the dashed-horizontal line. Comparisons within each individual plot correspond to comparisons of the estimation of β_{02} for differing values of β_{12} .

As can be seen in these individual plots, the variability in the estimation of the intercept parameter is heavily affected by the value of the slope parameter in the

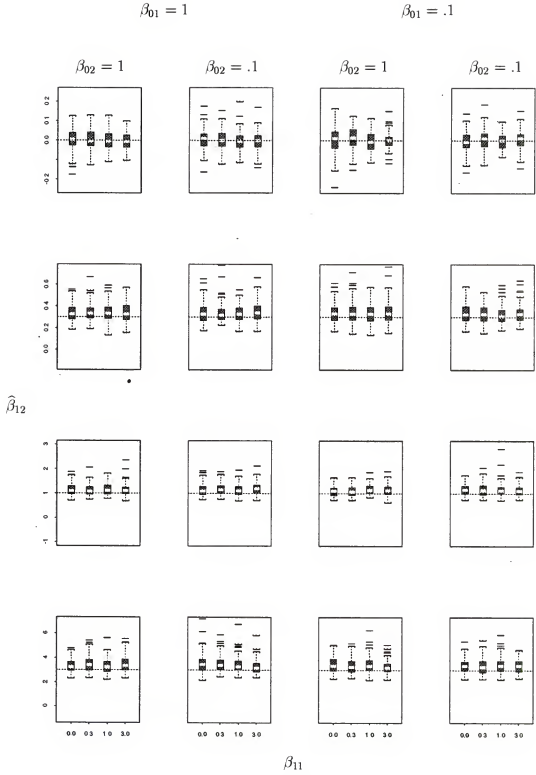


Figure 5.1. Estimation of the parameter β_{12} for sample size 20.

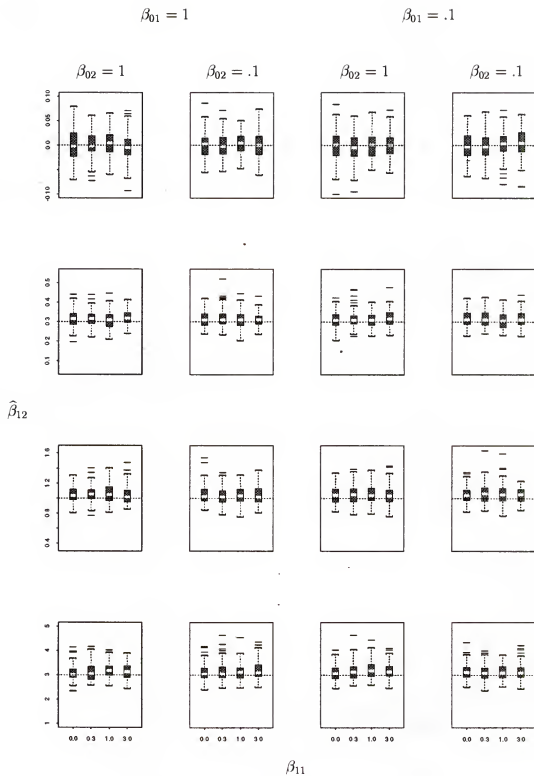


Figure 5.2. Estimation of the parameter β_{12} for sample size 50.

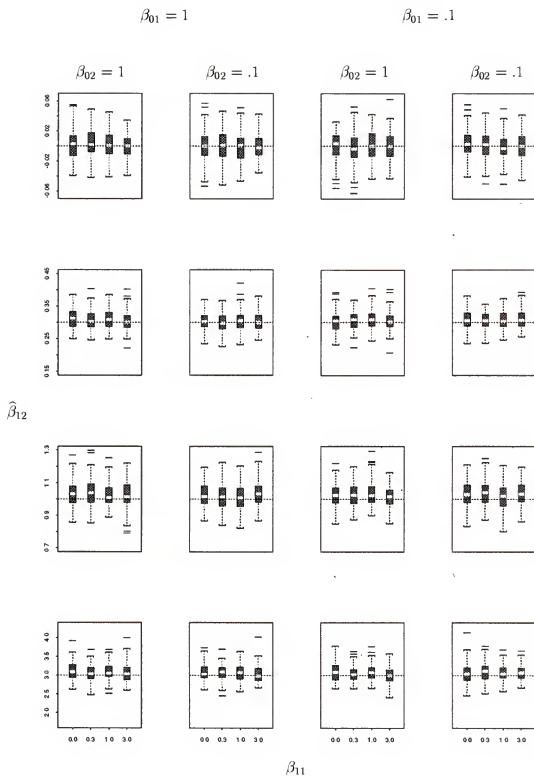


Figure 5.3. Estimation of the parameter β_{12} for sample size 100.

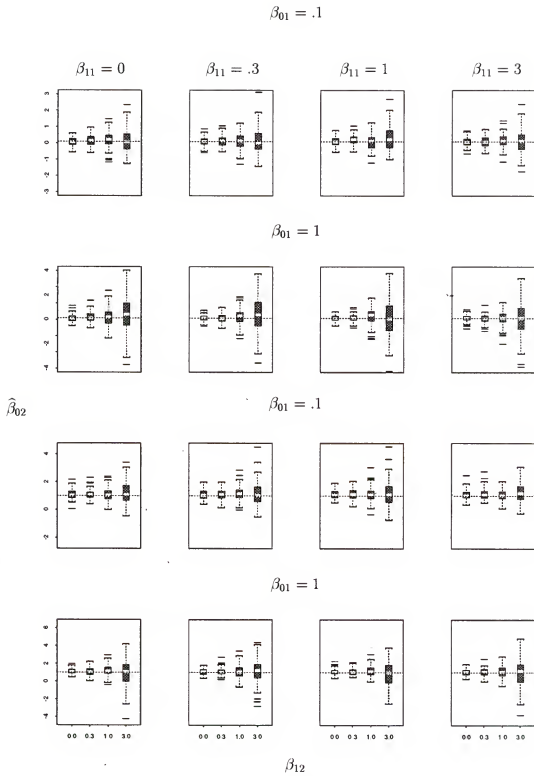
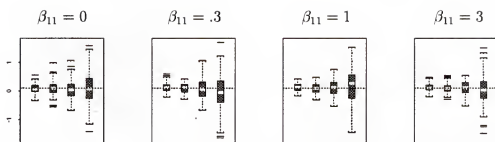
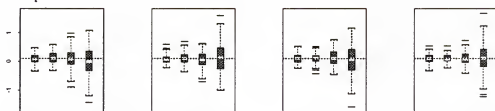


Figure 5.4. Estimation of the parameter β_{02} for sample size 20.

$$\beta_{01} = .1$$

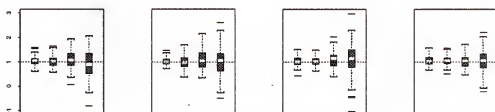


$$\beta_{01} = 1$$

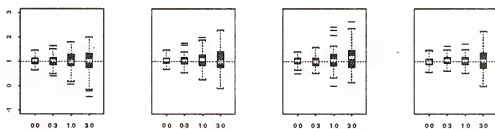


$$\hat{\beta}_{02}$$

$$\beta_{01} = .1$$



$$\beta_{01} = 1$$



$$\beta_{12}$$

Figure 5.5. Estimation of the parameter β_{12} for sample size 50.

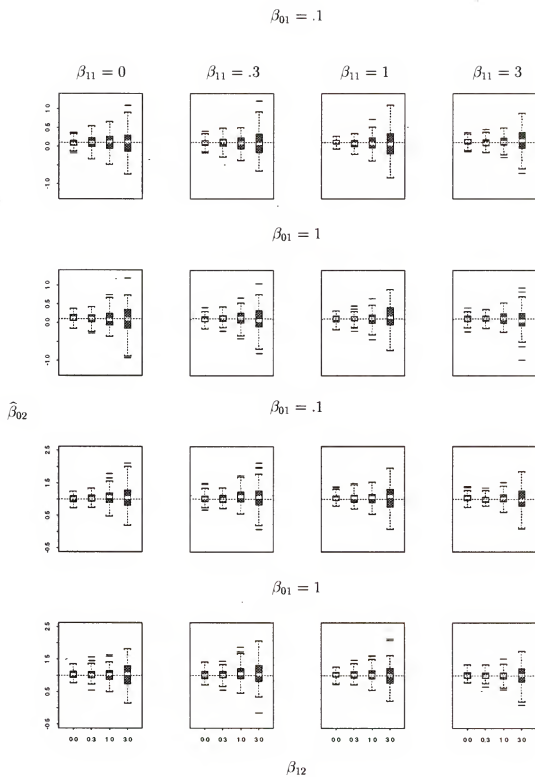


Figure 5.6. Estimation of the parameter β_{12} for sample size 100.

same term (β_{12}). Specifically, as the value of the slope term increases, the amount of variability in the estimated values of the intercept increases. This is not surprising, since this phenomenon also is found in the linear regression setting. Columns of this plot correspond to the estimation of an intercept parameter for differing values of the slope parameter in the other term (β_{11}). Little evidence of differences between columns is observed, indicating little dependence on the estimation of an intercept term on the other slope term. Neighboring rows correspond to estimation of one intercept while changing the other intercept. Little evidence of differences between neighboring rows implies little dependence in the estimation of an intercept for one term on the intercept for the other term.

Based on this small simulation study, it appears that the parameter estimates for the slope parameters are skewed and biased upwardly for small sample sizes. For samples of size one hundred (100), the slope estimates seem to be slightly biased; however, the skewness is no longer apparent. The intercept estimates do not seem to be appreciably skewed nor biased; however, the amount of variability in the estimates is affected by the values of the corresponding slope term.

5.3 Error Rates of Tests

The Wald and likelihood ratio tests are based on the asymptotic normality of the vector of maximum likelihood estimates. A small simulation study was performed to check the error rates of these two tests for small samples.

Five hundred data sets were generated for each combination of the following parameters:

$$n : \quad 20, 50, 100,$$

$$\beta_{01} : \quad 0, .1, .3, 1,$$

$$\beta_{02} : \quad 0, .1, .3, 1,$$

$$\begin{aligned}\beta_{11} : & \quad 0, .1, \\ \beta_{12} : & \quad 0, .1.\end{aligned}$$

The covariate values were once again generated as Uniform random variables on the interval $(-10,10)$, and held constant for all parameter choices. The null-hypothesis $H_0: (\beta_{11}, \beta_{12}) = (0, 0)$ was then tested for each combination of the parameters. The empirical Type I error rate and power were then calculated. Table 5.1 shows the empirical Type I error rate in relation to the nominal α level for a number of settings of the parameters. Table 5.2 shows the empirical power, computed for a number of different parameter settings.

It can be seen from Table 5.1 that the likelihood ratio test has a smaller (and better) Type I error rate than does the Wald test. The likelihood ratio test behaves quite well even for small sample sizes. For samples of size fifty or larger, the likelihood ratio test's empirical Type I error rate is quite close to the nominal α level. The Wald test's Type I error rate is consistently higher than the likelihood ratio's and also too liberal. Table 5.2 shows that, not unexpectedly, the Wald test has slightly greater power than the likelihood ratio test, although the two tests do not differ much. Due to the differences in Type I error rates between the two tests and the similarities of the powers for the two tests, it is recommended that the likelihood ratio test be used over the Wald test. This is especially true for small sample sizes.

5.4 Circular R^2 statistics

Four different circular R^2 measures were discussed in Chapter 4. The exact distributions of these statistics was not discussed due to their complicated form. The statistics were suggested as methods useful for determining the significance of terms in the model that contribute to the fit of the mean direction. It is hypothesized that

Table 5.1. Comparison of Type I error rates for two asymptotic tests.

					n=	20	n=	50	n=	100
α	β_{01}	β_{02}	β_{11}	β_{12}	LRT	Wald	LRT	Wald	LRT	Wald
.25	0.1	0.1	0	0	.290	.312	.260	.270	.276	.282
	0.1	0.3	0	0	.284	.314	.270	.302	.260	.270
	0.1	1.0	0	0	.302	.368	.250	.308	.260	.332
	0.3	0.1	0	0	.272	.304	.272	.290	.252	.268
	0.3	0.3	0	0	.282	.316	.284	.312	.254	.276
	0.3	1.0	0	0	.300	.374	.272	.336	.276	.344
	1.0	0.1	0	0	.306	.366	.284	.362	.254	.320
	1.0	0.3	0	0	.282	.372	.280	.350	.278	.368
	1.0	1.0	0	0	.296	.388	.248	.346	.250	.344
.10	0.1	0.1	0	0	.138	.158	.096	.110	.104	.108
	0.1	0.3	0	0	.142	.162	.110	.128	.124	.132
	0.1	1.0	0	0	.134	.190	.104	.162	.098	.142
	0.3	0.1	0	0	.140	.158	.104	.118	.094	.104
	0.3	0.3	0	0	.128	.142	.130	.146	.100	.106
	0.3	1.0	0	0	.128	.196	.114	.166	.112	.164
	1.0	0.1	0	0	.124	.188	.112	.184	.120	.176
	1.0	0.3	0	0	.128	.184	.124	.172	.132	.194
	1.0	1.0	0	0	.132	.204	.100	.158	.110	.180
.05	0.1	0.1	0	0	.074	.090	.054	.060	.068	.070
	0.1	0.3	0	0	.058	.086	.054	.062	.062	.078
	0.1	1.0	0	0	.080	.120	.062	.086	.056	.090
	0.3	0.1	0	0	.056	.082	.046	.052	.046	.048
	0.3	0.3	0	0	.082	.104	.072	.094	.042	.060
	0.3	1.0	0	0	.072	.106	.068	.100	.048	.084
	1.0	0.1	0	0	.072	.112	.060	.096	.064	.112
	1.0	0.3	0	0	.066	.108	.064	.112	.056	.112
	1.0	1.0	0	0	.074	.126	.052	.098	.054	.112

Table 5.2. Power estimates for two asymptotic tests.

					n=	20	n=	50	n=	100
α	β_{01}	β_{02}	β_{11}	β_{12}	LRT	Wald	LRT	Wald	LRT	Wald
.25	0.1	0.1	0	0.1	.844	.856	.992	.992	1.000	1.000
	0.1	0.3	0	0.1	.846	.860	.990	.990	1.000	1.000
	0.1	1.0	0	0.1	.818	.866	.976	.982	1.000	1.000
	0.3	0.1	0	0.1	.848	.862	.998	1.000	.998	.998
	0.3	0.3	0	0.1	.856	.874	.996	.996	1.000	1.000
	0.3	1.0	0	0.1	.820	.852	.988	.990	1.000	1.000
	1.0	0.1	0	0.1	.850	.886	.998	.998	1.000	1.000
	1.0	0.3	0	0.1	.876	.918	.988	.994	1.000	1.000
	1.0	1.0	0	0.1	.824	.874	.980	.994	1.000	1.000
.10	0.1	0.1	0	0.1	.682	.714	.956	.958	1.000	1.000
	0.1	0.3	0	0.1	.696	.704	.966	.968	.998	.998
	0.1	1.0	0	0.1	.650	.694	.948	.962	.998	.998
	0.3	0.1	0	0.1	.680	.728	.970	.980	.998	.998
	0.3	0.3	0	0.1	.714	.752	.982	.984	.998	.998
	0.3	1.0	0	0.1	.656	.716	.946	.968	.994	.996
	1.0	0.1	0	0.1	.714	.798	.984	.992	1.000	1.000
	1.0	0.3	0	0.1	.732	.798	.972	.978	1.000	1.000
	1.0	1.0	0	0.1	.660	.742	.952	.972	1.000	1.000
.05	0.1	0.1	0	0.1	.560	.592	.920	.924	1.000	1.000
	0.1	0.3	0	0.1	.588	.610	.946	.946	.998	.998
	0.1	1.0	0	0.1	.518	.588	.902	.922	.980	.990
	0.3	0.1	0	0.1	.556	.584	.936	.942	.994	.994
	0.3	0.3	0	0.1	.578	.628	.960	.966	.994	.996
	0.3	1.0	0	0.1	.532	.606	.892	.930	.992	.994
	1.0	0.1	0	0.1	.580	.682	.950	.980	.998	1.000
	1.0	0.3	0	0.1	.614	.714	.940	.968	.998	1.000
	1.0	1.0	0	0.1	.524	.628	.932	.956	.998	1.000

the statistics will increase along with the effect of the covariate on the mean direction. In order to test certain assumptions and properties of these statistics, a small simulation study was run.

One hundred data sets were generated for each combination of the following parameters:

$$\begin{aligned} n : & \quad 20, 50, 100, \\ \beta_{01} : & \quad .1, 1.0, \\ \beta_{02} : & \quad .1, 1.0, \\ \beta_{11} : & \quad 0, .3, 1.0, \\ \beta_{12} : & \quad 0, .3, 1.0. \end{aligned}$$

Once again the covariates were generated as Uniform random variables on the interval $(-10,1)$, and held constant for all parameter choices. Figures 5.7-5.24 give histograms of the different circular R^2 measures for differing values of the parameters. For the histograms, only one slope was allowed to be nonzero. The effect of allowing both slopes to be nonzero just magnified the effects seen in these histograms.

In general, all four circular R^2 measures are very close to zero when there is no covariate in the model, although there is a large amount of variability for small sample sizes and small concentrations. There were distinct shifts in location for all four statistics as the mean direction (and concentration) was allowed to depend on the covariate. In general, the R_{c2}^2 and R_{c4}^2 statistics (which are based on the log transformation) showed a larger location shift than did R_{c1}^2 and R_{c3}^2 . Unfortunately, these same statistics also showed more variability when the covariate was not in the model.

$$\beta_{01} = \beta_{02} = .1, \beta_{11} = 0$$

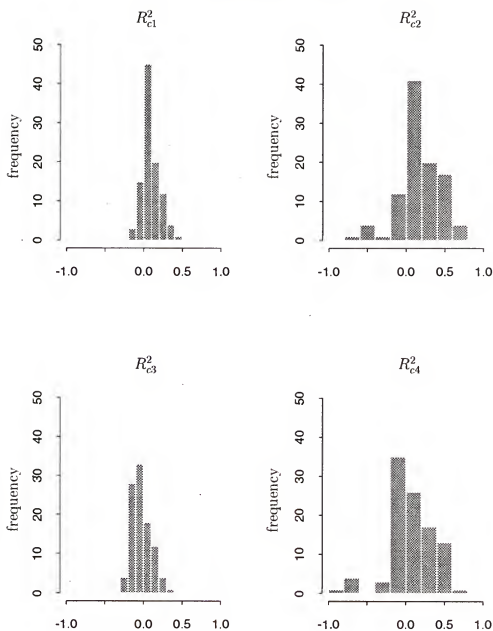


Figure 5.7. Empirical distribution of the circular R^2 values, $n=20$.

$$\beta_{01} = \beta_{02} = .1, \beta_{11} = 0.3$$

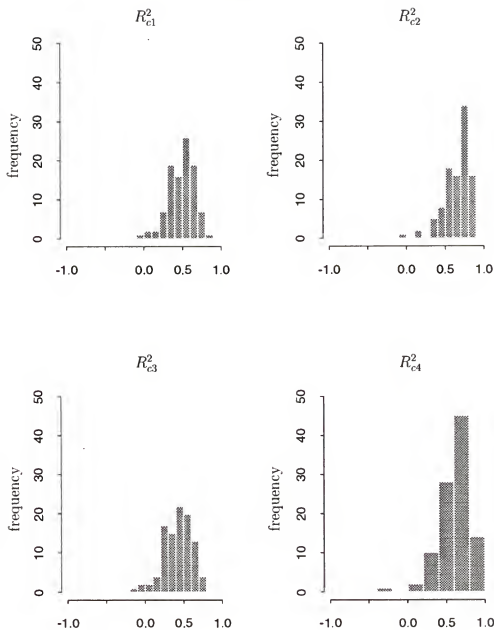


Figure 5.8. Empirical distribution of the circular R^2 values, $n=20$.

$$\beta_{01} = \beta_{02} = .1, \beta_{11} = 1.0$$

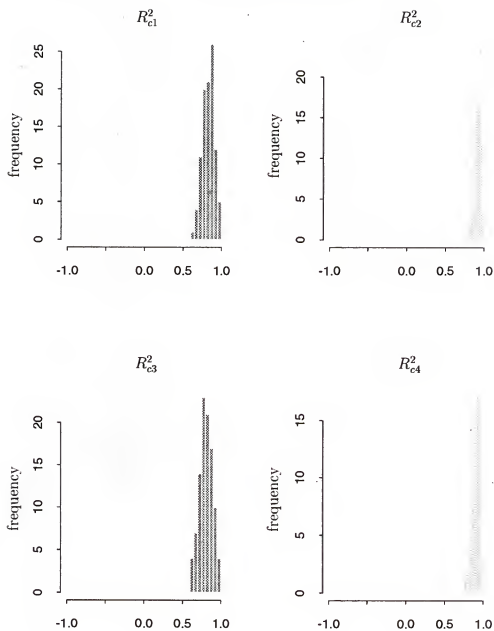


Figure 5.9. Empirical distribution of the circular R^2 values, $n=20$.

$$\beta_{01} = \beta_{02} = 1.0, \beta_{11} = 0$$

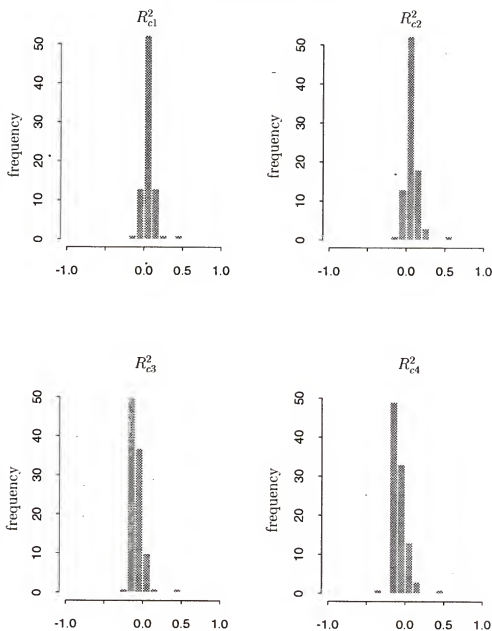


Figure 5.10. Empirical distribution of the circular R^2 values, $n=20$.

$$\beta_{01} = \beta_{02} = 1.0, \beta_{11} = 0.3$$

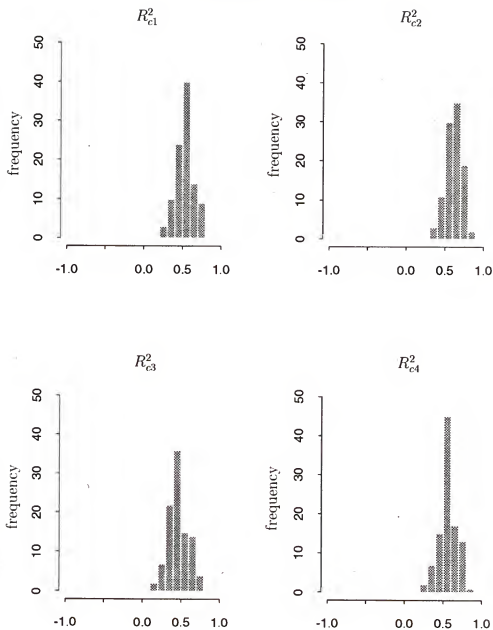


Figure 5.11. Empirical distribution of the circular R^2 values, $n=20$.

$$\beta_{01} = \beta_{02} = 1.0, \beta_{11} = 1.0$$

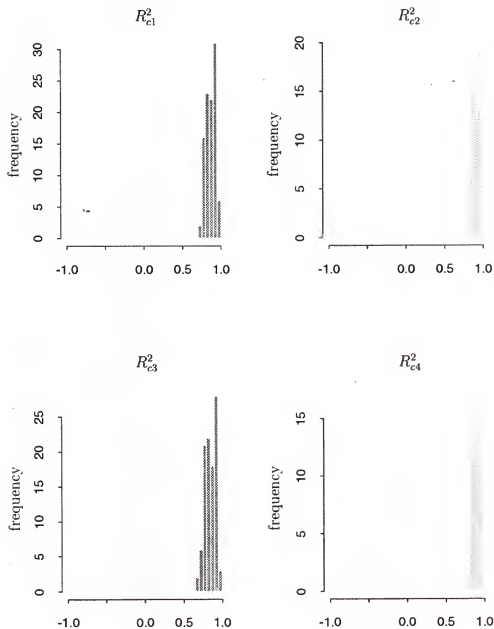


Figure 5.12. Empirical distribution of the circular R^2 values, $n=20$.

$$\beta_{01} = \beta_{02} = .1, \beta_{11} = 0$$

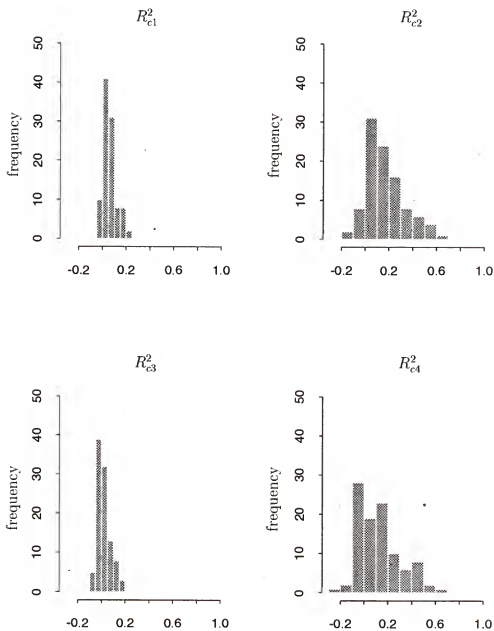


Figure 5.13. Empirical distribution of the circular R^2 values, $n=50$.

$$\beta_{01} = \beta_{02} = .1, \beta_{11} = 0.3$$

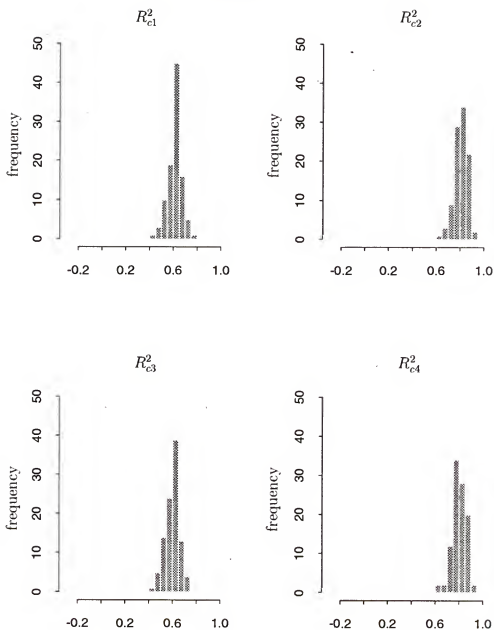


Figure 5.14. Empirical distribution of the circular R^2 values, $n=50$.

$$\beta_{01} = \beta_{02} = .1, \beta_{11} = 1.0$$

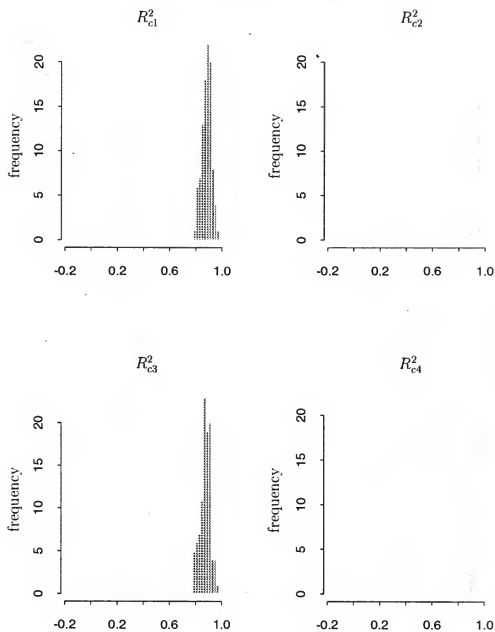


Figure 5.15. Empirical distribution of the circular R^2 values, $n=50$.

$$\beta_{01} = \beta_{02} = 1.0, \beta_{11} = 0$$

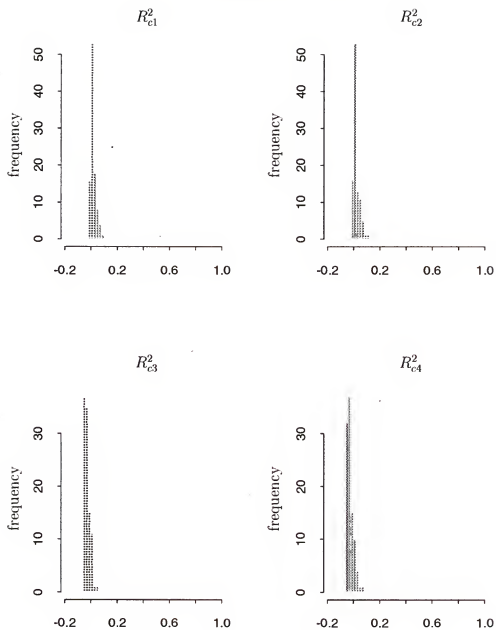


Figure 5.16. Empirical distribution of the circular R^2 values, $n=50$.

$$\beta_{01} = \beta_{02} = 1.0, \beta_{11} = 0.3$$

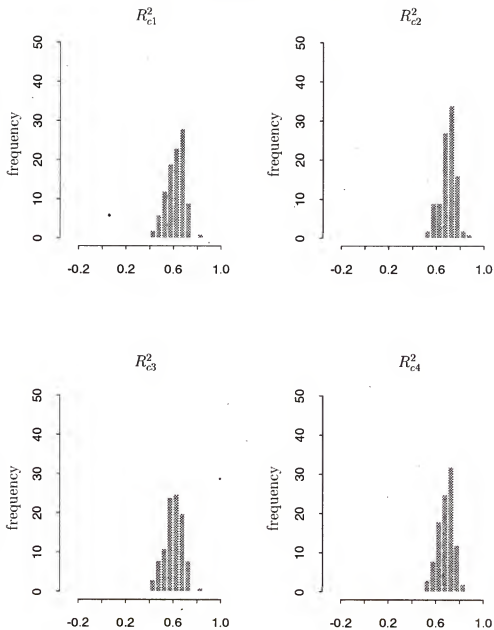


Figure 5.17. Empirical distribution of the circular R^2 values, $n=50$.

$$\beta_{01} = \beta_{02} = 1.0, \beta_{11} = 1.0$$

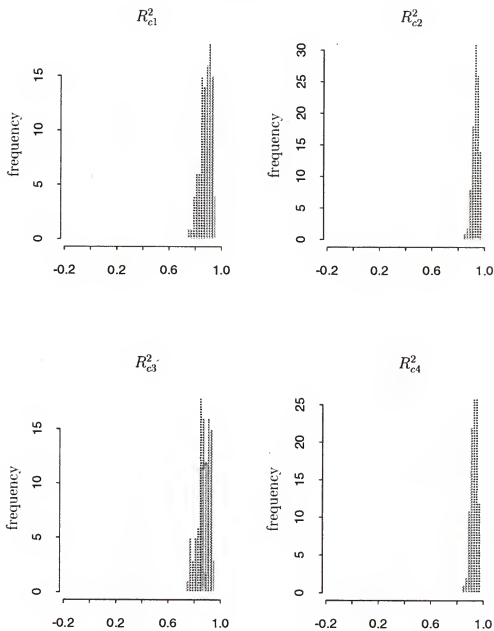


Figure 5.18. Empirical distribution of the circular R^2 values, $n=50$.

$$\beta_{01} = \beta_{02} = .1, \beta_{11} = 0$$

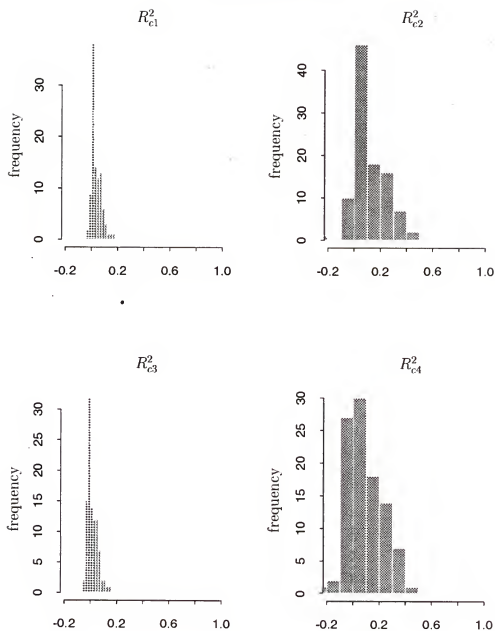


Figure 5.19. Empirical distribution of the circular R^2 values, $n=100$.

$$\beta_{01} = \beta_{02} = .1, \beta_{11} = 0.3$$

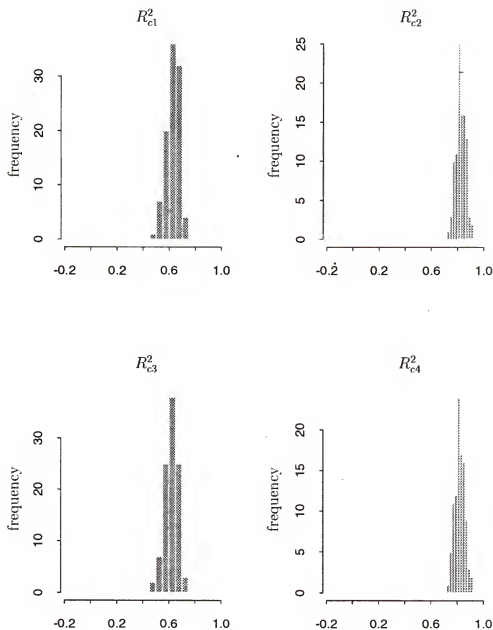


Figure 5.20. Empirical distribution of the circular R^2 values, $n=100$.

$$\beta_{01} = \beta_{02} = .1, \beta_{11} = 1.0$$

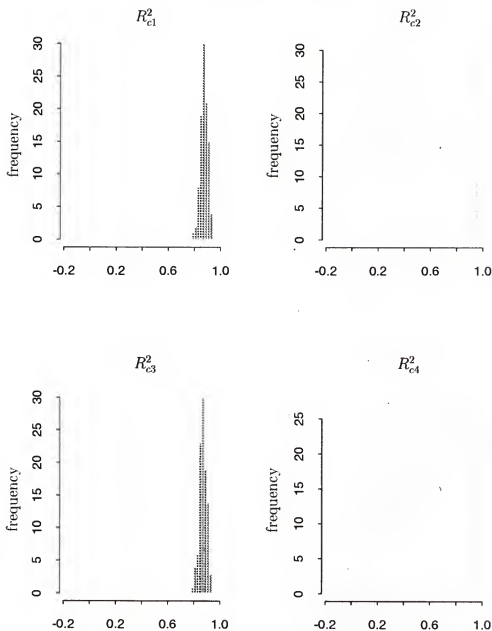


Figure 5.21. Empirical distribution of the circular R^2 values, $n=100$.

$$\beta_{01} = \beta_{02} = 1.0, \beta_{11} = 0$$

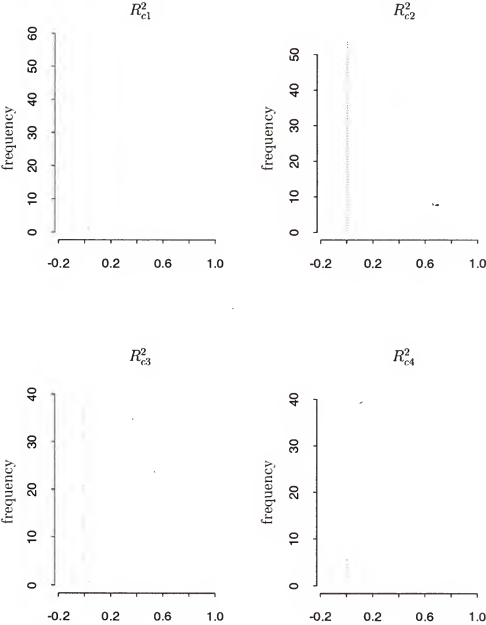


Figure 5.22. Empirical distribution of the circular R^2 values, n=100.

$$\beta_{01} = \beta_{02} = 1.0, \beta_{11} = 0.3$$

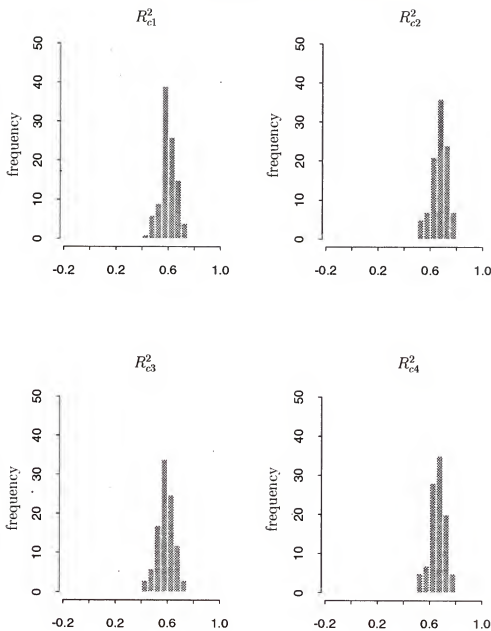


Figure 5.23. Empirical distribution of the circular R^2 values, $n=100$.

$$\beta_{01} = \beta_{02} = 1.0, \beta_{11} = 1.0$$

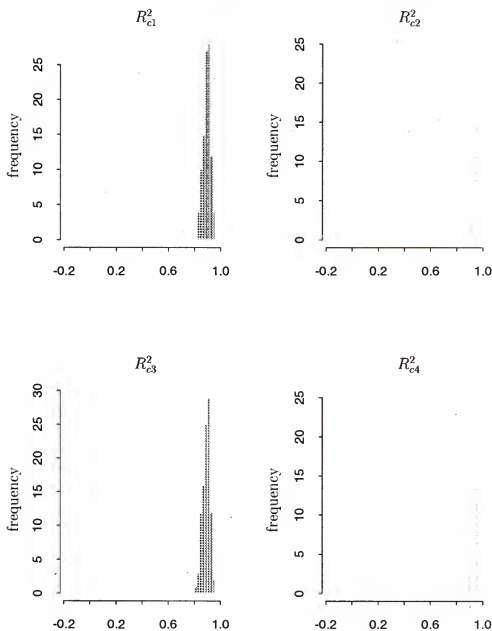


Figure 5.24. Empirical distribution of the circular R^2 values, $n=100$.

CHAPTER 6 EXAMPLES

6.1 Introduction

In this chapter, data from four different studies are analyzed using the methods discussed in this dissertation. Three of the examples involve data from ecological studies which measured directions of movement and covariates expected to influence these directions. The first such example relates the direction of travel of displaced blue periwinkles to distance moved. The second considers the possible influence of on-shore light sources on the direction of travel of newly hatched loggerhead turtles. The third example is from an industrial setting, and involves the balancing of automotive flywheels. Finally, the fourth study involves migratory patterns of three species of butterfly.

6.2 Fisher and Lee (1992) Periwinkle Data

Fisher and Lee apply their models to data on the directional movement of small blue periwinkles (*Nodilittorina Unifasciata*). The data were generated through a series of experiments discussed in Underwood and Chapman (1985, 1989) and are replicated in Table 6.1 with the directions transformed to radians. The measurements are of the directions and distances moved by small blue periwinkles after they were transported downshore from the area where they normally lived. Data are plotted in Figure 6.1, with the Y -axis corresponding to direction. Due to the cyclic nature of directions, the top and bottom of this plot are actually quite close together. Based on the plotted

Table 6.1. Fisher and Lee (1992) blue periwinkle data.

Distance in cm (x)	107	46	33	67	122	69
Direction(θ)	1.6937	1.1519	1.2915	1.0647	1.0123	1.0472
Distance in cm (x)	43	30	12	25	37	69
Direction(θ)	1.7453	1.5533	2.9845	2.8972	.7104	1.0472
Distance in cm (x)	5	83	68	38	21	1
Direction(θ)	3.4383	1.7104	1.5010	2.1468	2.8798	2.3213
Distance in cm (x)	71	60	71	71	57	53
Direction(θ)	1.7628	1.8326	1.2392	1.4661	1.3090	1.7104
Distance in cm (x)	38	70	7	48	7	21
Direction(θ)	1.4486	1.2392	1.2915	1.5882	0.6632	3.4907
Distance in cm (x)	27					
Direction(θ)	.9774					

data there seems to be some relationship between the direction of movement of the periwinkle and the distance moved. In particular, judging by the data, there seems to be some evidence that both the mean direction and concentration depend on distance moved.

Table 6.2. Analysis of the Fisher and Lee data using the new model.

Factors	$-2 \log L$	LRT(P-value)	R^2 's	
Dist, Dist ²	38.8279		$R_{c1}^2 = .2007$ $R_{c3}^2 = .0728$	$R_{c2}^2 = .2225$ $R_{c4}^2 = .0981$
Dist	41.1996	2.3717(.3054)	$R_{c1}^2 = .1948$ $R_{c3}^2 = .1351$	$R_{c2}^2 = .2161$ $R_{c4}^2 = .1580$
Intercept	66.6381	25.4385(0.0)	--	

The new model was fit to this data and Table 6.2 shows the analysis and comparisons of three nested models. The highest order model fits quadratic terms, and the simplest model fits a constant mean direction and concentration. The value of the likelihood ratio test (LRT) along with the P-value give the significance of the parameters for next higher order model. In other words, these tests are comparing the reduced model to the next higher order full model. The parameter tests imply that the best fitting model is the model that includes only linear terms. The circular R^2

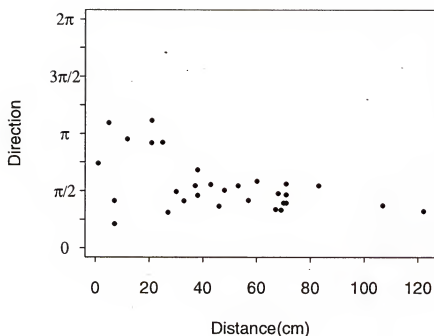


Figure 6.1. Fisher and Lee (1992) data.

values back this up, by implying that including quadratic terms doesn't improve the mean fit. The linear terms are a significant improvement over the constant mean and concentration model. This can be seen through both the likelihood ratio tests and the circular R^2 values. The circular R^2 values imply that allowing this single covariate into the model explains between fifteen and twenty percent of the variability in the data. The parameter estimates and their standard errors are shown in the following table.

Parameter	Estimate	Std. Error
β_{01}	-1.1228	.3515
β_{11}	0.0298	.0069
β_{02}	0.1573	.3846
β_{12}	0.0489	.0010

The resulting fitted model gives

$$\begin{aligned}\hat{\mu}_x &= \tan^{-1}(-1.1228 + 0.0298x, 0.1573 + 0.0489x), \\ \hat{\kappa}_x^* &= \left[(-1.1228 + 0.0298x)^2 + (0.1573 + 0.0489x)^2\right]^{\frac{1}{2}} \\ &= \left[1.2854 - 0.0516x + 0.0033x^2\right]^{\frac{1}{2}}.\end{aligned}$$

Fisher and Lee recognized the differing concentrations through the data and suggested their mixed model for analysis based on the von Mises distribution. Their fitted model corresponds to

$$\begin{aligned}\hat{\mu}_x &= 2.0438 + 2 \tan^{-1}(-0.009(x - 47.6516)), \\ \hat{\kappa}_x &= \exp(1.78 + 0.045(x - 47.6516)).\end{aligned}$$

As a part of the estimation for the mixed model parameters, Fisher and Lee fit a mean only model that corresponds to

$$\begin{aligned}\hat{\mu}_x &= 1.693 + 2 \tan^{-1}(-0.013(x - 47.6516)), \\ \hat{\kappa} &= 3.2.\end{aligned}$$

Figure 6.2 compares the fits of the new model with the fits of the Fisher and Lee models for the Periwinkle data. Over the range of the data, the new model fits as well as, if not better than, the models proposed by Fisher and Lee. The Fisher and Lee mean model fits the data quite well although the assumption of constant concentration does not seem to hold. As an additional comparison between the new model and those of Fisher and Lee, the circular R^2 values for the two models were calculated and compared to the circular R^2 values of our new model. The circular R^2 values for the Fisher and Lee models can be found in the following table along with the values for the new model. Based on these comparisons and Figure 6.2, neither the Fisher and Lee mixed model, which fits the mean direction worse than a constant

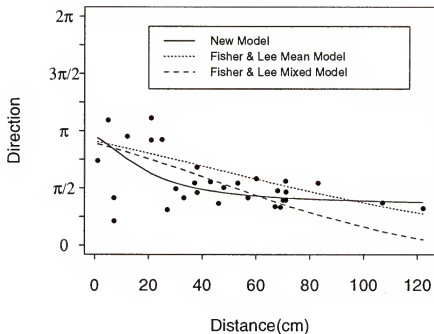


Figure 6.2. Comparison of three models.

mean direction, nor the Fisher and Lee mean model do as good a job as the new model in fitting the mean direction. In terms of the amount of variability explained, the new model explains close to three times as much variability as the Fisher and Lee mean model.

R^2	FL mean model	FL mixed model	new model
R_{c1}^2	.0627	-.0027	.1948
R_{c2}^2	.0708	-.0031	.2161
R_{c3}^2	.0304	-.0373	.1351
R_{c4}^2	.0388	-.0377	.1580

Confidence bands and prediction bands about the mean direction will allow us to compare the fit of the model with the underlying concentration of the data. It is straightforward to calculate the confidence bands for the mean direction, since the

estimated variance covariance matrix of $(\beta_{01}, \beta_{11}, \beta_{02}, \beta_{12})'$ is

$$\begin{bmatrix} .123519 & -.002063 & -.001891 & -.000464 \\ -.002063 & .000048 & -.000237 & .000022 \\ -.001891 & -.000237 & .147891 & -.002858 \\ -.000464 & .000022 & -.002858 & .000100 \end{bmatrix}$$

Figure 6.3 shows the mean fit along with a set of approximate pointwise 95% confidence bands for the mean. As expected, the confidence bands are wider when the covariate is small. This seems natural since when a periwinkle has not traveled far there is little useful information about its direction.

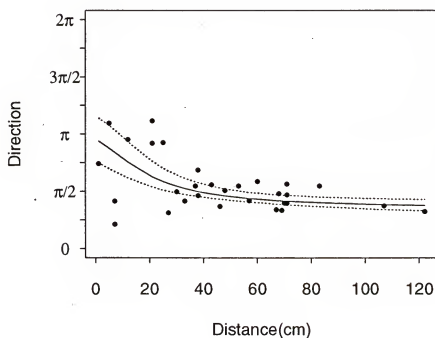


Figure 6.3. Approximate 95% pointwise confidence bands for the mean direction.

Finally, prediction bands are constructed for our fitted model. Two separate techniques are used to construct prediction bands for a single future observation. The

first technique is the Monte-Carlo method discussed in Section 4.4.5 that is symmetric about the mean direction. For these prediction bands, ten thousand (10,000) observations were simulated at each covariate value. Their corresponding directions were recorded and the endpoints of the symmetric interval which contains the mean direction and 95% of the observations were then recorded. The upper and lower bands were then estimated by using the Lowess function in S-plus, to give smooth bands. The simpler method of equating the offset normal to the von Mises distribution and then to the normal distribution was also used. Figure 6.4 gives both sets of prediction bands. As expected, the technique using the normal approximation to the offset normal distribution is narrower than the Monte-Carlo technique when the covariate is small. This is due to the fact that, for small covariates, the associated concentration values are relatively small and the normal approximation is valid only for large concentrations. The Monte-Carlo method should be valid for all concentrations as long as the sample size of the original data set is relatively large so that the normal approximation of the parameter estimates is valid. It is good to note that when the covariate, and hence the concentration, is large the two techniques are almost indistinguishable.

6.3 Light Intensity and Turtle Hatchling Data

A project involving the study of turtle hatchlings was conducted by Sheila Colwell of the Florida Cooperative Fish and Wildlife Research Unit. It has been shown that hatchling sea turtles depend on visual cues to locate the ocean after emerging from nests at night. Sea-finding failure is rare, except in instances where hatchlings emerge on beaches illuminated with anthropogenic lights. These bright lights dominate the radiance of surrounding directions. In those instances, hatchlings ignore the visual cues of surrounding directions, and orient toward the artificial light source.

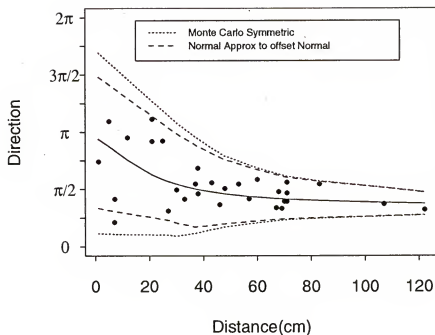


Figure 6.4. Comparison of two approximate 95% prediction bands for a future observation.

Data were collected over a span of seven nights, during the 1994 calendar year, at three different locations, 13A, 11, and 13 in the Eglin Air Force Base, which is located in the Panhandle of the state of Florida. These locations were chosen, because of the extremely bright lights shining onto the beach area from Eglin Air Force Base. The sample nights were chosen based on the readiness of loggerhead hatchlings to emerge from their nest. A circular area was defined at each site by a string boundary with a diameter of six meters. Intervals of $5^\circ(\frac{\pi}{36})$ width were marked on the string boundary by their medians. The zero direction was chosen as the point on the circle closest to the water. The hatchlings were collected just prior to emergence and small groups of turtles were then released from the center of

Table 6.3. Sheila Colwell's turtle hatchling data.

	13				11		13A	
	9/1	10/4	10/12	10/27	9/27	10/25	9/28	cont.
1	3.5779	3.7524	2.0071	2.7925	4.0142	3.8397	3.8397	3.3161
2	3.2288	2.5307	2.1816	2.9670	4.1887	4.3633	4.0142	2.7052
3	2.6179	1.8325	3.4033	3.1415	4.2760	4.5378	4.3633	3.7524
4	2.5307	2.0071	4.1887	3.1415	4.3633	4.6251	4.4505	4.1887
5	2.1816	3.8397	2.3561	3.8397	4.7123	4.8869	5.1487	4.1887
6	3.5779	3.6651	2.5307	4.0142	4.2760	3.4033	3.4906	4.3633
7	2.4434	3.5779	2.7052	3.2288	4.4505	3.8397	4.2760	4.1887
8	2.4434	3.4033	3.2288	3.4906	4.4505	4.2760	4.2760	4.4505
9	2.4434		3.2288	3.5779	4.7996	4.2760	4.3633	5.0614
10	2.4434		4.5378	3.9269	3.4906	4.7996	5.1487	4.3633
11	3.8397		2.4434	4.4505	4.5378	4.2760	5.3232	
12	3.5779		2.7052		4.7123	4.6251	3.4906	
13	2.9670		2.7925		3.4906	4.6251	3.8397	
14	2.4434		3.1415		4.1015	5.4977	3.8397	
15	2.3561		3.9269		4.1887	3.1415	5.0614	
16	3.0543		3.4033		4.5378	4.7123	5.3232	
17	2.7925		2.0943		4.6251	4.6251	5.6723	
18	2.4434		2.7925		4.1887	4.7996	4.9741	
19	1.8325		3.7524		3.6651	3.5779	4.6251	
20	1.6580		4.6251		4.1015	4.4505	3.9269	
21	3.7524				4.2760	4.3633	3.8397	
22	2.5307					4.7123	3.6651	
23	2.5307					2.7925	3.6651	
24	2.5307						3.6651	
25	2.4434						3.7524	
26	4.1887						3.9269	
27	3.6651						4.2760	
28	3.0543						4.5378	
29	2.4434						3.6651	
30	3.5779						3.8397	
31	3.2288						4.3633	
32	2.3561						4.6251	
33	1.7453						4.7123	
34	3.4906						4.3633	
35	3.1415						3.3161	
36	2.6179						3.6651	
37	2.3561						3.9269	
38	2.1816						4.1015	
39	2.0943						4.2760	
40	3.4906						4.2760	

fit to this data with night nested within location. Table 6.4 shows the analysis and comparisons of the different models fit using the nested design. The highest degree model is the full nested model and the simplest model is the model that assumes constant mean and concentration throughout all nights for each individual location.

Table 6.4. Nested ANOVA analysis of the turtle data using our new model.

Factors	$-2 \log L$	LRT	R^2 's	
Loc, Night(Loc)	304.9961		$R^2_{c1} = .5146$ $R^2_{c3} = .4788$	$R^2_{c2} = .5648$ $R^2_{c4} = .5318$
Loc, Night(Loc=13)	308.1000	3.1039(.2118)	$R^2_{c1} = .5134$ $R^2_{c3} = .4830$	$R^2_{c2} = .5636$ $R^2_{c4} = .5363$
Loc	322.9189	17.9228(.0218)	$R^2_{c1} = .4786$ $R^2_{c3} = .4661$	$R^2_{c2} = .5292$ $R^2_{c4} = .5179$
Intercept	464.7161	141.7972(0.0)	—	

Both the tests and the circular R^2 values imply that the most parsimonious model is a model whose location effects are different and where there is no night effect within location 11 but there is a night effect for location 13. This model fits the data quite well. Figure 6.6 shows the fitted means along with approximate 95% prediction bands using the method of equating the offset normal to the von Mises and then to the normal distribution. This approximation should be quite close since the data are very concentrated.

Unfortunately this model does not give us any information relating the directional travel of the turtles to the light intensity. In order to include information on light intensity we have to include the measured light intensity measurements. Two separate filters, A and B, were used to measure the radiance, but only filter A will be discussed here. Unfortunately, each directional observation was not influenced by a single radiance so an appropriate covariate is not obvious. Because each turtle was influenced by all light intensities in all directions, it would seem reasonable to use some statistic, calculated from the light intensity measurements, as the covariate.

Figure 6.7 shows the plotted light intensity for location 13 on 9/1. This plot shows

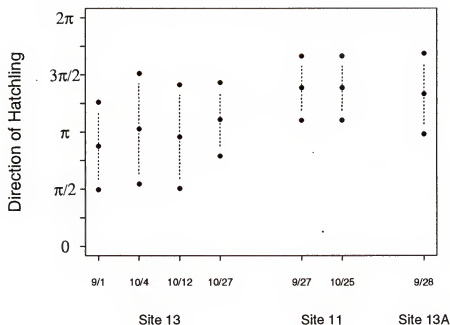


Figure 6.6. Fitted values and prediction bands for the turtle data with a nested design.

the magnitude of the light intensity towards certain extremely bright lights around Eglin Air Force Base in relation to the radiance reflected off the Gulf of Mexico. Due to the shape of the plot, and others like it, a natural covariate would be the mean direction of light intensity. An estimate of the direction of brightest light intensity can be found through a weighted average of the observed directions weighted by the measured light intensity, i.e. define the mean direction of light intensity (x_{mli}) as

$$x_{mli} = \tan^{-1} \left(\sum_{i=1}^{24} LI_i * \sin \left(\frac{i\pi}{12} \right), \sum_{i=1}^{24} LI_i * \cos \left(\frac{i\pi}{12} \right) \right),$$

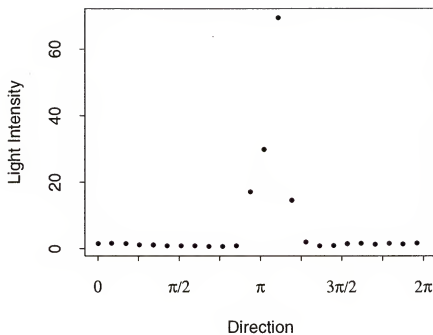


Figure 6.7. Plotted light intensities versus angle for location 13 on 9/1.

where LI_i corresponds to the light intensity for the direction defined by $\frac{i\pi}{12}$. These estimated mean light intensities are in fact, quite close to the mean direction in which the turtles traveled, and both are listed below.

Location	Night	x_{mli}	$\hat{\mu}_{turt}$
13	9/1	3.3364	2.7822
	10/4	3.2488	3.1334
	10/12	3.3735	3.0584
	10/27	3.4049	3.4994
11	9/27	3.8598	4.2651
	10/25	3.6966	4.3896
13A	9/28	3.6234	4.1998

The data were refit using the estimated mean direction of light intensity as the covariate. At first glance it may not seem that a linear function of a direction makes sense. However, the range of the estimated mean direction of light intensity is in a very narrow interval, so that linear functions are reasonable over the range of the data. Table 6.3 shows that the most parsimonious model, based on both the circular

Table 6.5. Regression analysis of turtle data using the new model.

Factors	$-2 \log L$	LRT(P-value)	R^2 's	
MLI, MLI ²	324.7961		$R_{c1}^2 = .4902$ $R_{c3}^2 = .4780$	$R_{c2}^2 = .5408$ $R_{c4}^2 = .5297$
MLI	329.0862	4.2901(.1171)	$R_{c1}^2 = .4761$ $R_{c3}^2 = .4699$	$R_{c2}^2 = .5267$ $R_{c4}^2 = .5211$
Intercept	464.7161	135.6299(0.0)	--	

R^2 values and the likelihood ratio tests, corresponds to the model with only the singular covariate MLI. This model improves a great deal over the constant mean and concentration model, and explains about fifty percent of the variability in the data. The best fitting regression model and the best fitting nested ANOVA model are very comparable in terms of their fits. The circular R^2 values are very comparable, and the log-likelihood attains a very similar magnitude. The best fitting nested ANOVA model contains ten distinct parameters, whereas the regression model using linear terms contains only four parameters. The parameter estimates and their standard errors are found in the following table.

Parameter	Estimate	Std. Error
β_{01}	-4.7829	1.4619
β_{11}	1.0081	0.4085
β_{02}	20.6969	1.6901
β_{12}	-6.1255	0.4836

The resulting fitted model gives

$$\hat{\mu}_x = \tan^{-1}(-4.7829 + 1.0081x, 20.6969 - 6.1255x),$$

$$\begin{aligned}\hat{\kappa}_x^* &= \left[(-4.7829 + 1.0081x)^2 + (20.6969 - 6.1255x)^2 \right]^{\frac{1}{2}} \\ &= \left[451.2378 - 263.2010x + 38.5380x^2 \right]^{\frac{1}{2}}.\end{aligned}$$

Figure 6.8 shows the plotted data along with approximate pointwise 95% confidence bands and two sets of approximate 95% prediction bands for the turtle hatchling direction. The simple method of equating the offset normal to the normal distribution is used as well as the Monte Carlo method with nonsymmetric confidence bands. These Monte Carlo bands are based on one thousand (1,000) simulated observations for each instance of the covariate.

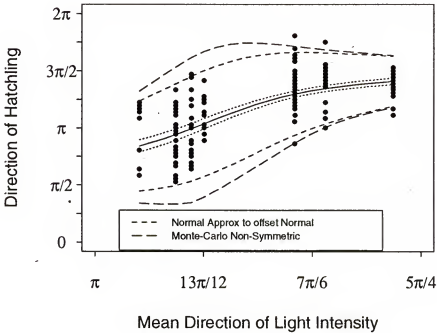


Figure 6.8. Confidence and prediction bands for the turtle data.

6.4 Anderson and Wu (1994) Automotive Flywheel Data

Anderson and Wu (1994a, 1995b) analyzed a data set concerning the balancing of automotive flywheels. The data resulted from a 2^4 factorial design with ten (10) replicates, where the response was the direction, or angle, at which a corrective weight should be placed on the flywheel in order to balance it. The four factors that were of interest, were

- A the location of a butt weld on the flywheel, at a fixed(F) or random(R) point on the flywheel.
- B the flywheel radial grade, Low(L) or High(H).
- C the thickness grade of the flywheel, Low(L) or High(H).
- D the size of the Counter-weight, Low(L) or High(H).

The data are not replicated here due to space.

Anderson and Wu analyze the data in the two papers for two separate purposes. The data are fit with a concentration only model assuming constant mean in Anderson and Wu (1994a) and with a mean only model assuming constant concentration in Anderson and Wu (1994b). The new model simultaneously models both the concentration and the location. It is of interest to compare our final model and its corresponding interpretations against their two separate models. Anderson and Wu concluded that the factors A, C, and D and the interactions AD and CD were affecting the mean direction, and that the factors B, C, and D and the interactions BC and CD were affecting the concentration.

Table 6.6 gives the results of comparisons of the main effects model, the second order model, the third order model and the full factorial model. Based on table 6.6, it seems that the most appropriate model may include three factor interactions. To get a better feel for this, each factor was systematically deleted from the full model,

Table 6.6. Preliminary analysis of the Anderson and Wu (1994) data using the new model.

Factors	$-2 \log L$	LRT(P-value)	R^2 's	
Full Model	404.8879		$R^2_{c1} = .2427$ $R^2_{c3} = .0652$	$R^2_{c2} = .3247$ $R^2_{c4} = .1667$
Third Order Model	405.1062	.2183(.8966)	$R^2_{c1} = .2425$ $R^2_{c3} = .0793$	$R^2_{c2} = .3245$ $R^2_{c4} = .1790$
Second Order Model	420.2265	15.1203(.0568)	$R^2_{c1} = .2058$ $R^2_{c3} = .0907$	$R^2_{c2} = .2805$ $R^2_{c4} = .1762$
First Order Model	470.4805	50.254(0.0)	$R^2_{c1} = .0671$ $R^2_{c3} = .0173$	$R^2_{c2} = .0985$ $R^2_{c4} = .0504$
Constant	514.1680	43.6875(0.0)	--	

and the effects of the factors were ranked in order as done by Anderson and Wu. Table 6.7 shows the results of these tests, along with the associated rankings for the Anderson and Wu concentration (using the log link function) and mean models. Judging by the rankings of each term and their associated P-values, it looks as though only two three-way interactions, may be important. These correspond to the three-way interactions ABC and ACD. Also, only three two-way interactions seem even marginally significant. A new model was fit including all main effects and the first-order interactions AD, BC, and CD as well as the three-way interactions ABC and ACD. The ABC interaction term was then retested using this reduced model and found not significant at $\alpha = .10$, resulting in a final model that included all main effects and the interactions AD, BC, CD and ACD.

One final set of tests is necessary. Each factor and all of its associated interactions are systematically removed from the final model, and the corresponding values of the circular R^2 are compared. This will allow us to test a factors influence on the mean direction and the concentration. Table 6.8 shows the results of these tests. Judging by the R^2 values, it seems that the terms B and BC are only affecting the fit of the concentration. Deleting them does not seem to affect the R^2 terms. The other three

Table 6.7. Ordering of factors.

Source	New Model		Anderson and Wu Mean Model		Anderson and Wu Conc. Model	
	P-value	Rank	P-value	Rank	P-value	Rank
A	.1902	(9)	.284	(7)	—	(12)
B	.0025	(5)	.943	(15)	—	(1)
C	.0002	(2)	.069	(3)	—	(4)
D	.0010	(4)	.565	(11)	—	(5)
AB	.3293	(10)	.696	(12)	—	(10)
AC	.4679	(14)	.327	(8)	—	(15)
AD	.0010	(3)	.034	(2)	—	(7)
BC	.0154	(6)	.120	(4)	—	(3)
BD	.4166	(12)	.550	(10)	—	(11)
CD	.0000	(1)	.017	(1)	—	(2)
ABC	.0754	(8)	.190	(6)	—	(6)
ABD	.4421	(13)	.820	(14)	—	(8)
ACD	.0398	(7)	.124	(5)	—	(9)
BCD	.3776	(11)	.339	(9)	—	(14)
ABCD	.8966	(15)	.791	(13)	—	(13)

factors seem, at least in part, to be affecting the mean direction. This conclusion is similar to the conclusions given by Anderson and Wu in their two papers.

Anderson and Wu argue that it would be most economical if the mean direction was close to π with a concentration as large as possible. To decide on the best settings of the factors, Table 6.9 gives the estimated mean directions and concentrations based on our final model. Unfortunately, the best choices of mean direction and concentration (denoted by *) do not coincide with one another. Thus it would be up to the plant manager to choose which response is more important. The plant manager would most likely choose between the set $(A, B, C, D) = \{(R, L, L, L), (R, H, H, L), (R, H, H, H), (F, H, H, H)\}$ dependent on costs.

Table 6.8. Factor comparisons for the final model.

Factors	$-2 \log L$	LRT(P-value)	R^2 's	
A,B,C,D,AD,BC, CD, ACD	419.0594	—	$R^2_{c1} = .2132$ $R^2_{c3} = .1245$	$R^2_{c2} = .2894$ $R^2_{c4} = .2094$
No A: B,C,D,BC,CD	443.3519	24.2925(.0005)	$R^2_{c1} = .1277$ $R^2_{c3} = .0688$	$R^2_{c2} = .1813$ $R^2_{c4} = .1260$
No B: A,C,D,AD,CD,ACD	439.6778	20.6184(.0004)	$R^2_{c1} = .2062$ $R^2_{c3} = .1410$	$R^2_{c2} = .2810$ $R^2_{c4} = .2219$
No C: A,B,D,AD	475.5958	56.5364(0.0)	$R^2_{c1} = -.1358$ $R^2_{c3} = -.1963$	$R^2_{c2} = -.2264$ $R^2_{c4} = -.2918$
No D: A,B,C,BC	480.8341	61.7747(0.0)	$R^2_{c1} = .0292$ $R^2_{c3} = -.0226$	$R^2_{c2} = .0438$ $R^2_{c4} = -.0722$

6.5 Butterfly Data

Walker and Littell discuss a large observational study involving the migratory patterns of three species of butterfly, Gulf Fritillary (*Argaulis vanillae*), Long-Tailed Skipper (*Urbanus proteus*), and Cloudless Sulphur (*Phoebis sennae*), as they migrate into the state of Florida each fall. Walker and Littell attempted to describe and analyze the migratory directions of these butterflies and relate the directional flight patterns to a number of covariates. One question that was not fully answered by the original study involved the relationship between the directional patterns of the butterflies and geographical location. Specifically, on the Atlantic coast, there seemed to be evidence that the butterflies were “sensing” the coast before reaching the coastline and adjusting their migratory patterns appropriately. In order to help better answer this question, a more extensive study was done by Walker and J. J. Whitesell. Data was collected on a transect extending from the Atlantic coast at O’Neil, Florida west-erly to Valdosta, Georgia over four years. Vanishing angles of the three species of butterflies were recorded along with numerous covariates at a number of sites along the transect. The actual data corresponded to more than two thousand five hundred observations, the majority of which were for the Cloudless Sulphur species. Figures

Table 6.9. Estimated mean directions and concentrations for the final model.

A	B	C	D	$\hat{\mu}$	$\hat{\kappa}^*$
R	L	L	L	3.0500	1.7000
R	L	L	H	0.7693	0.4693
R	L	H	L	2.7135	1.0194
R	L	H	H	3.0906*	0.5254
R	H	L	L	2.9098	1.5961
R	H	L	H	0.8456	0.7185
R	H	H	L	2.8016	2.5569
R	H	H	H	2.9182	2.0590
F	L	L	L	2.3859	1.4804
F	L	L	H	5.6456	0.2710
F	L	H	L	1.6053	0.3644
F	L	H	H	2.6900	1.0487
F	H	L	L	2.2238	1.5441
F	H	L	H	0.1387	0.3606
F	H	H	L	2.6537	1.6932
F	H	H	H	2.7911	2.5838*

6.9-6.11 show the plotted directions on the Y -axis along with the distance from the coast on the X -axis for the three species of butterfly.

Based on the previous work of Walker and Littell and some plotting of the data, the following factors were thought to be very influential in the directional patterns of these butterflies.

- Distance from coast in km (Dist).
- Squared distance from coast in km (Dist²).
- Observer effect (Obs).
- Time of day (TOD).

Only these factors were included as possible factors in the model.

Table 6.10 gives comparisons of several nested models. The results of Table 6.10 imply that distance from coast is a significant factor in predicting the directional travel of the Long-Tailed Skipper as it migrates into the state of Florida. There is

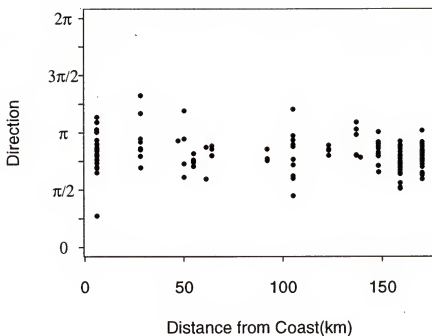


Figure 6.9. Plot of the Long Tailed Skipper data.

also slight evidence that there is an observer effect in the prediction of Long-Tailed Skipper directions. However, neither factor seems to be having much of an affect on the prediction of the mean direction. This can be noted by the extremely low values of the R^2 terms for all the models. The best fitting models only explain between two and four percent of the overall variability in the data. This seems to imply that the factors distance form coast and observer are only affecting the concentration of the directions. This is further backed up by the fitted model with only distance from coast included in the model. The fitted mean direction differs only slightly from a horizontal line. The fitted mean direction along with 95% confidence and prediction bands for the mean direction are shown in Figure 6.12. The prediction bands are

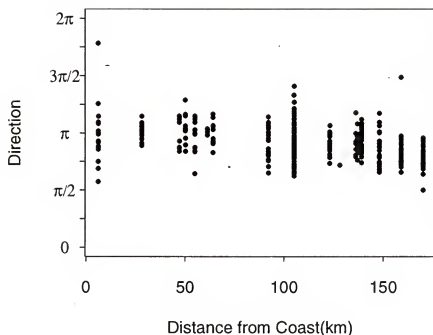


Figure 6.10. Plot of the Gulf Fritillary data.

calculated using the normal approximation to the von Mises and then equating the von Mises distribution to the offset normal distribution.

The new model was then fit to the Gulf Fritillary data. The results of the fits along with tests of parameters appear in Table 6.11 and imply that both distance from coast and observer are significant factors in predicting the directional travel of the Gulf Fritillary as it migrates into the state of Florida. For the Gulf Fritillary, there seems to be a large effect of the distance from coast on the mean direction. This can be noted by the large change in R^2 as this factor is introduced into the model. The effect due to distance from coast in the model accounts for approximately 15% of the overall variability in the data. The effect of observer, once again does not seem to influence the fit of the mean direction, only the fitted concentration. The

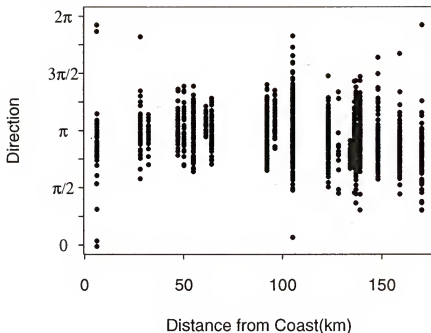


Figure 6.11. Plot of the Cloudless Sulphur data.

fitted mean direction along with 95% confidence and prediction bands for the mean direction are shown in Figure 6.13 when the model includes only distance from coast as the covariate. The prediction bands are calculated using the normal approximation to the von Mises and then equating the von Mises distribution to the offset normal distribution.

Finally the new model was fit to the Cloudless Sulphur data. The results of the fits along with tests of parameters appear in Table 6.12. Table 6.12 implies that all the factors in the model, distance from coast, squared distance from coast, observer, and time of day, are all significant factors in predicting the directional travel of the Cloudless Sulphur as it migrates into the state of Florida. For the Cloudless Sulphur, there seems to be a large effect of the distance from coast on the mean directional

Table 6.10. Analysis of the Long-Tailed Skipper data.

Factors	$-2 \log L$	LRT(P-value)	R^2 's	
Dist, Dist^2 , Obs, TOD	142.5512		$R_{c1}^2 = .0382$ $R_{c3}^2 = -.1519$	$R_{c2}^2 = .0399$ $R_{c4}^2 = -.1343$
Dist, Dist^2 , Obs	142.5795	.0283(.9859)	$R_{c1}^2 = .0393$ $R_{c3}^2 = .0002$	$R_{c2}^2 = .0414$ $R_{c4}^2 = .0020$
Dist, Obs	145.9003	3.3208(.1901)	$R_{c1}^2 = .0424$ $R_{c3}^2 = .0166$	$R_{c2}^2 = .0429$ $R_{c4}^2 = .0185$
Dist	150.7761	4.8758(.0873)	$R_{c1}^2 = .0236$ $R_{c3}^2 = .0105$	$R_{c2}^2 = .0246$ $R_{c4}^2 = .0116$
Intercept Only	161.6568	10.8881(.0043)	--	

Table 6.11. Analysis of the Gulf Fritillary data.

Factors	$-2 \log L$	LRT(P-value)	R^2 's	
Dist, Dist^2 , Obs, TOD	466.4950		$R_{c1}^2 = .1547$ $R_{c3}^2 = .1403$	$R_{c2}^2 = .1611$ $R_{c4}^2 = .1468$
Dist, Dist^2 , Obs	468.9904	2.4954(.2872)	$R_{c1}^2 = .1493$ $R_{c3}^2 = .1384$	$R_{c2}^2 = .1556$ $R_{c4}^2 = .1448$
Dist, Obs	472.6587	3.6683(.1597)	$R_{c1}^2 = .1483$ $R_{c3}^2 = .1411$	$R_{c2}^2 = .1546$ $R_{c4}^2 = .1474$
Dist	478.7598	6.1011(.0473)	$R_{c1}^2 = .1458$ $R_{c3}^2 = .1422$	$R_{c2}^2 = .1519$ $R_{c4}^2 = .1484$
Intercept Only	571.6009	92.8411(0.0)	--	

travel of the butterfly. In this situation, it is necessary to include the squared distance from the coast also, as it is heavily influencing the mean direction. The effects due to distance from coast and squared distance from coast account for approximately 15% of the overall variability in the data. The effects of observer and time of day do not seem to influence the fit of the mean direction, only the fitted concentration. The fitted mean direction along with 95% confidence and prediction bands for the mean direction are shown in Figure 6.14 when the model includes only distance from coast and squared distance from coast as the covariates. This model corresponds to the returning model discussed in Section 4.2.3. The prediction bands are calculated

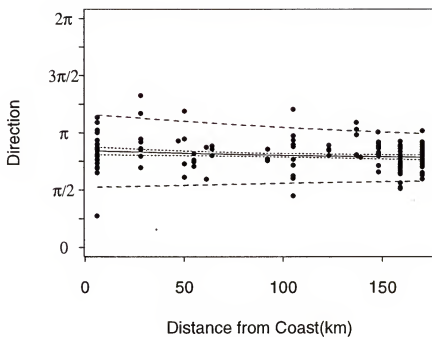


Figure 6.12. Fitted model along with 95% confidence and prediction bands for the Long-Tailed Skipper data.

using the normal approximation to the von Mises and then equating the von Mises distribution to the offset normal distribution.

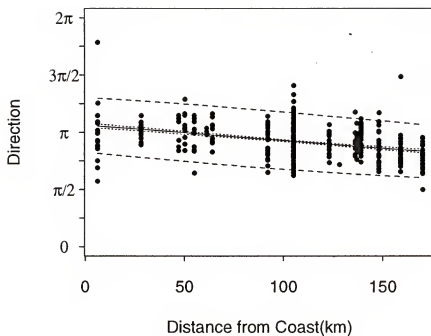


Figure 6.13. Fitted model along with 95% confidence and prediction bands for the Gulf Fritillary data.

Table 6.12. Analysis of the Cloudless Sulphur data.

Factors	$-2 \log L$	LRT(P-value)	R^2 's	
Dist, Dist ² , Obs, TOD	2926.6690		$R_{c1}^2 = .1559$ $R_{c3}^2 = .1522$	$R_{c2}^2 = .1683$ $R_{c4}^2 = .1646$
Dist, Dist ² , TOD	2940.8577	14.1887(.0008)	$R_{c1}^2 = .1510$ $R_{c3}^2 = .1482$	$R_{c2}^2 = .1630$ $R_{c4}^2 = .1602$
Dist, Dist ²	2972.6357	31.7780(0.0)	$R_{c1}^2 = .1443$ $R_{c3}^2 = .1424$	$R_{c2}^2 = .1558$ $R_{c4}^2 = .1540$
Dist	3166.6301	193.9944(0.0)	$R_{c1}^2 = .0731$ $R_{c3}^2 = .0721$	$R_{c2}^2 = .0795$ $R_{c4}^2 = .0785$
Intercept Only	3332.1826	165.5525(0.0)	--	

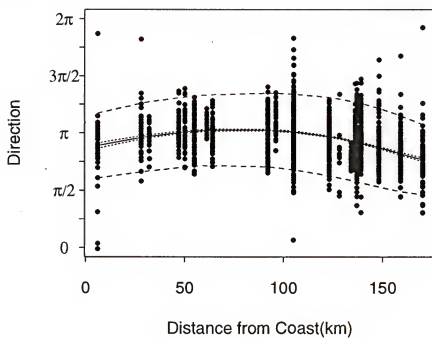


Figure 6.14. Fitted model along with 95% confidence and prediction bands for the Cloudless Sulphur data.

CHAPTER 7 CONCLUSIONS AND FUTURE RESEARCH

7.1 Conclusions

This dissertation begins with a discussion of the offset normal distribution, a under-utilized distribution for directional data. Distributional properties are discussed and the distribution is compared to the standard distribution used for inference about directional data. A regression model based on this distribution is then introduced that is applicable in a wide variety of settings.

The new regression model requires both the mean direction and the concentration parameter of an offset normal random variable to be functions of a set of covariates or regressor variables. The new regression model has an obvious estimation technique based on the EM algorithm and the uniqueness of the parameter estimates is guaranteed. These parameter estimates correspond to maximum likelihood estimates. Standard tests of parameters based on the asymptotic distribution of the maximum likelihood estimates can be applied for large sample sizes. Confidence and prediction bands for the estimated mean direction and/or the estimated concentration parameter can be calculated using the delta method. The model seems to work quite well in comparison to the other methods found in the literature, and allows for the covariates to be either dummy or continuous variables.

Goodness of fit statistics for the new model are introduced that behave in a similar fashion to the R^2 and adjusted R^2 statistics for scalar responses. These measures, defined as circular R^2 's, measure the goodness of fit in terms of the mean direction

only. These measures allow for detecting the influence of factors which only affect either the mean direction or the concentration.

A simulation study then looks at the distributional properties of the parameter estimates, and the appropriateness of the asymptotic properties of these estimates. This simulation also checks the appropriateness and error rates of some suggested tests. Simulation studies also are run to infer distributional properties of the goodness of fit measures, since simple distributional forms are not possible.

Finally the new model is applied to a number of different data sets to show the wide range of applications possessed by the model. The model is seen to fit the data sets quite well and give further insight into other models. The new model also has other advantages over existing models. Estimation difficulties are not encountered for the new model, that were known to exist for models in the literature that limited the application of the existing models to data with small numbers of covariates.

7.2 Future Research

One area for future research is to extend the present model to allow for non-symmetric, or even bimodal distributions of the directional response. This complication can be allowed into the current model by modifying the underlying variance of the imagined bivariate normal distribution, i.e. replacing \mathbf{I} in equation 4.1 by a general matrix Σ . Some structure to this matrix can be assumed or it can be allowed to vary freely. The idea behind estimation for the resulting model will not change. However, the resulting distribution of the directional observation will change, and the estimation of the parameters will become more complicated. It is not known if this generalization will lead to identifiability problems. The present model can be described as a special case of this more general model.

A separate area for future research involves the use of a link function to relate the linear function of the covariates to the underlying bivariate normal mean. The mean

vector of the bivariate normal in equation 4.1 could be replaced by the mean vector

$$\begin{pmatrix} h_1(\mathbf{x}'_i/\beta_1) \\ h_2(\mathbf{x}'_i/\beta_2) \end{pmatrix},$$

where h_1 and h_2 are link functions chosen by the experimenter. One possible situation where a model of this form would be appropriate corresponds to axial data. Axial data are directional data that essentially lie on only the half circle. Due to this constraint, it would seem natural that the bivariate normal distribution that is generating the directional observations to have a mean on only half of the two dimensional plane. Choosing one of the link functions to be a monotone mapping of the real line onto the positive reals would then satisfy this restriction, e.g. one of the h 's could be chosen as $\exp(\cdot)$.

An additional area for future research concerns the use of nonparametric smoothing techniques. Several methods of smoothing using local regression can be found in the literature for linear responses. Similar techniques could be used for directional responses based on this model. The Loess method of Cleveland (1979) is one such method where the new regression model can be incorporated. The new regression model would be applied locally and then with the addition of certain continuity restrictions, the local regressions would be smoothed together. One complication with this technique is the large number of observations necessary for such a technique. The new regression model, in general, requires twice as many parameters as a regression model for linear data, and hence would be very costly in the total number of parameters required.

A simple and obvious extension of the new regression model discussed in this dissertation concerns spherical data. Regression models for spherical responses have similar difficulties to those with circular responses. The formulation behind the new regression model does not restrict the direction to be circular. The same arguments can be applied for spherical data. In the spherical regression setting the directions on

the sphere would correspond to the directions of trivariate normal random variables. The EM algorithm would then be used with slight modifications to iterate between the complete data specification and the incomplete data specification.

Finally other situations in which complications for regression models have been found could benefit from the approach behind the new model. Specifically, by modeling observed data as being incomplete and being the projections of some higher order data. One area where this technique could be applied concerns compositional data, where the response can be thought of as occurring on the surface of a simplex.

APPENDIX A TRIGONOMETRIC MOMENTS OF THE OFFSET NORMAL DISTRIBUTION

This appendix gives some fundamental arguments concerning trigonometric moments, the appropriate moments for directional data. The specific form of the trigonometric moments for the offset normal distribution are also found. Much of this material follows the notation and arguments of Mardia (p. 41, 44, and 62).

The trigonometric moments are defined by

$$\alpha_p = \mathbf{E}(\cos(p\theta)) \quad \text{and} \quad \beta_p = \mathbf{E}(\sin(p\theta)).$$

The simplest method of finding the trigonometric moments is through the use of the characteristic function. The characteristic function takes the form

$$\phi_p = \mathbf{E}\{\exp(ip\theta)\} = \mathbf{E}(\cos(p\theta) + i\sin(p\theta)), \quad p = 0, \pm 1, \pm 2, \dots$$

hence

$$\phi_p = \alpha_p + i\beta_p. \tag{A.1}$$

Since the characteristic function can be viewed as a function in the complex plane, equation A.1 can be written as

$$\phi_p = \rho_p \exp^{i\mu_p}.$$

The p^{th} central moments (about μ) correspond to

$$\bar{\alpha}_p = \mathbf{E}(\cos(p(\theta - \mu))) \quad \text{and} \quad \bar{\beta}_p = \mathbf{E}(\sin(p(\theta - \mu))),$$

which can be written as

$$\bar{\alpha}_p = \rho_p \cos(\mu_p - p\mu) \quad \text{and} \quad \bar{\beta}_p = \rho_p \sin(\mu_p - p\mu),$$

where $\mu_1 = \mu$. Because of the relationship between the central and non-central moments, it is sufficient to assume that $\mu = 0$.

For the offset normal distribution, taking $\mu = 0$ and concentration parameter $\kappa \geq 0$ gives

$$\beta_p = \mathbf{E}(\sin p\theta) = 0,$$

for all p . Therefore

$$\phi_p = \alpha_p,$$

and in particular

$$\phi_0 = \mathbf{E}(\cos 0) = 1.$$

Due to the cosine function being even, it can be assumed that p is positive, i.e.

$$\phi_{-p} = \phi_p.$$

For $p \neq 0$

$$\begin{aligned} \phi_p &= \alpha_p \\ &= \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{\sqrt{2\pi}} \int_0^{2\pi} \kappa \cos \theta \cos(p\theta) \exp\left(\frac{1}{2}\kappa^2 \cos^2(\theta)\right) \Phi(\kappa \cos \theta) d\theta. \end{aligned}$$

Expanding the exponential term in its Taylor series and then expanding Φ in its Taylor Series gives

$$\begin{aligned} \phi_p &= \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{\sqrt{2\pi}} \int_0^{2\pi} \sum_{i=0}^{\infty} \frac{(\kappa \cos \theta)^{2i+1}}{2^i i!} \cos(p\theta) \Phi(\kappa \cos \theta) d\theta \\ &= \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{\sqrt{2\pi}} \sum_{i=0}^{\infty} \frac{\kappa^{2i+1}}{2^i i!} \int_0^{2\pi} (\cos \theta)^{2i+1} \cos(p\theta) \Phi(\kappa \cos \theta) d\theta \\ &= \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{\sqrt{2\pi}} \sum_{i=0}^{\infty} \frac{\kappa^{2i+1}}{2^i i!} \int_0^{2\pi} (\cos \theta)^{2i+1} \cos(p\theta) \left[\frac{1}{2} + \frac{\kappa \cos \theta}{\sqrt{2\pi}} + \right. \\ &\quad \left. \frac{1}{\sqrt{2\pi}} \sum_{n=2}^{\infty} \frac{(\kappa \cos \theta)^{2n-1} (-1)^{n+1}}{(2n-1)(2n-2)2^{n-2}(n-2)!} \right] d\theta \end{aligned}$$

$$\begin{aligned}
&= \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{2\sqrt{2\pi}} \sum_{i=0}^{\infty} \frac{\kappa^{2i+1}}{2^i i!} \int_0^{2\pi} (\cos \theta)^{2i+1} \cos(p\theta) d\theta + \\
&\quad \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{2\pi} \sum_{i=0}^{\infty} \frac{\kappa^{2i+2}}{2^i i!} \int_0^{2\pi} (\cos \theta)^{2i+2} \cos(p\theta) d\theta + \\
&\quad \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{2\pi} \sum_{i=0}^{\infty} \frac{\kappa^{2i}}{2^i i!} \left[\sum_{n=2}^{\infty} \frac{(-1)^{n+1} \kappa^{2n-1}}{(2n-1)(2n-2)2^{n-2}(n-2)!} \times \right. \\
&\quad \left. \int_0^{2\pi} (\cos \theta)^{2i+2n} \cos(p\theta) d\theta \right].
\end{aligned}$$

Following the notation of Mardia (p. 62), this can be written as

$$\begin{aligned}
\phi_p &= \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{2\sqrt{2\pi}} \sum_{i=0}^{\infty} \frac{\kappa^{2i+1}}{2^i i!} T_{(2i+1,p)} + \\
&\quad \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{2\pi} \sum_{i=0}^{\infty} \frac{\kappa^{2i+2}}{2^i i!} T_{(2i+2,p)} + \\
&\quad \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{2\pi} \sum_{i=0}^{\infty} \frac{\kappa^{2i}}{2^i i!} \left[\sum_{n=2}^{\infty} \frac{(-1)^{n+1} \kappa^{2n-1}}{(2n-1)(2n-2)2^{n-2}(n-2)!} T_{(2i+2n,p)} \right].
\end{aligned}$$

This equation can now be simplified, since the function $T_{(a,b)}$ is zero whenever its arguments do not have the same parity (are not both even or both odd). Hence for p odd, the last two terms on the right hand side are zero yielding

$$\phi_p = \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{2\sqrt{2\pi}} \sum_{i=0}^{\infty} \frac{\kappa^{2i+1}}{2^i i!} T_{(2i+1,p)}.$$

Mardia gives a general solution of for $T_{a,b}$ of the form

$$T_{2r+p,p} = \frac{2\pi(2r+p)!}{r!(r+p)!2^{2r+p}}, \quad r > 0. \quad (\text{A.2})$$

Plugging this into the above equation gives for p odd

$$\phi_p = \frac{\sqrt{2\pi} \exp\left(-\frac{1}{2}\kappa^2\right)}{4} \sum_{i=(p-1)/2}^{\infty} \frac{\kappa^{2i+1} (2i+1)!}{2^{3i} i! (i + \frac{1-p}{2})! (i + \frac{p+1}{2})!}.$$

For $p = 1$ this corresponds to

$$\phi_1 = \frac{\sqrt{2\pi} \exp\left(-\frac{1}{2}\kappa^2\right)}{4} \sum_{i=0}^{\infty} \frac{\kappa^{2i+1}(2i+1)!}{2^{3i}(i!)^2(i+1)!},$$

which is equivalent to equation 2.11 in the text.

For p even, the first term on the right hand side is zero, giving

$$\begin{aligned} \phi_p &= \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{2\pi} \sum_{i=0}^{\infty} \frac{\kappa^{2i+2}}{2^i i!} T_{(2i+2,p)} + \\ &\quad \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{2\pi} \sum_{i=0}^{\infty} \frac{\kappa^{2i}}{2^i i!} \left[\sum_{n=2}^{\infty} \frac{(-1)^{n+1} \kappa^{2n-1}}{(2n-1)(2n-2)2^{n-2}(n-2)!} T_{(2i+2n,p)} \right] \\ &= \frac{\exp\left(-\frac{1}{2}\kappa^2\right)}{4} \sum_{i=(p-2)/2}^{\infty} \frac{\kappa^{2i+2}(2i+2)!}{2^{3i} i! (i+1-\frac{p}{2})! (i+1+\frac{p}{2})!} + 4 \exp\left(-\frac{1}{2}\kappa^2\right) \sum_{i=0}^{\infty} \frac{\kappa^{2i-1}}{2^{3i} i!} \times \\ &\quad \left[\sum_{n=\max(2,p/(2-i))}^{\infty} \frac{(-1)^{n+1} \kappa^{2n-1} (2i+2n)!}{(2n-1)(2n-2)2^{3n}(n-2)! (i+n-\frac{p}{2})! (i+n+\frac{p}{2})!} \right]. \end{aligned}$$

All of these terms seem to be related to Bessel functions; however, no simplified form has been found for these equations.

APPENDIX B

ESTIMATING THE PARAMETERS OF THE FISHER AND LEE MEAN MODEL

The likelihood of the Fisher and Lee mean model is multimodal. Typically there are at least $3 \times \text{dimension}(\beta)$ local maxima on the likelihood surface. Because of this, Fisher and Lee suggested an iterative scoring algorithm along with graphical exploration of the likelihood to uniquely estimate the regression parameters. However, this is only feasible for one or two covariates.

The likelihood is multimodal due to the form of the model. Because the model must travel around the circle exactly once, the height of the likelihood is very similar when an element of β , β_i , is zero and when that same element tends toward $\pm\infty$. The two models will only differ when the associated elements of x_{ij} are very close to zero. In fact if no observations occur at $x = 0$, the likelihood will achieve the same height at $\beta_i = 0$ and as $\beta_i \rightarrow \infty$. This problem, by itself, makes it such that there will be at least $2 \times \text{dimension}(\beta)$ maxima on the likelihood surface.

Obviously, as the dimensionality of the covariates increases, the resulting likelihood can have many local maxima. For small sample sizes and a large number of covariates, there will be a number of models that fit equally well. Even with graphical exploration of the log-likelihood, a suitable starting value for the iterative procedure may not be obvious. Because of these problems a different method of finding the global maximum seems necessary.

The problem of maximizing the likelihood can be improved by the use of computational methods developed for solving multimodal functions. One such algorithm is due to Price who suggested an algorithm to solve a multimodal function that requires

only evaluation of the function itself. Price's algorithm can be defined in the following steps:

- Select a "store" of starting points in the parameter space. (This store is typically chosen as uniform over a closed parameter space.)
- Calculate the value of the function at all points in the store.
- Generate a new point by reflecting a randomly chosen point in the store through centroids of a subset of other points in the store.
- Exchange new point for the worst point in the store, if new point is better than the worst.
- Continue this process until all points in store are sufficiently close together.

There is really only a single parameter that must be chosen by the experimenter, the size of the store. Both Price and Conlon (1992) suggest that a linear function of the dimension of the parameter space worked well. Specifically, Conlon suggested that the size of the store be on the order of $10 \times (\dim \mathbf{x}_i)$ or $20 \times (\dim \mathbf{x}_i)$, where $\dim \mathbf{x}_i$ is the dimension of the covariate vector or the number of covariates in the model.

Price's algorithm, as stated above, did not work sufficiently well even for very large choices of the store. A modification was made based on knowledge of the likelihoods. In one dimension the difficulties with the likelihood tend to occur when the global maximum is very close to zero. When the global maximum is away from zero the likelihood is better behaved, and Price's procedure, as described, worked well. It would seem reasonable to create the store with this in mind. Generalizing this to higher dimensions, the modification could be implemented as follows.

- Choose half of the store locally around the origin. (This could be done in a number of ways, for the simulation studies the values were chosen as uniform over the hyper-square $(-.01, .01)$.)

- Choose the other half of the store uniformly throughout a larger closed space, that contains all reasonable values of the parameters. (To get a feel for what can be considered as reasonable values of the parameters, the standardized data were analyzed.)

A small simulation study was conducted to test the effectiveness of the modification of Price's algorithm. von Mises random variables were generated using the algorithm of Best and Fisher (1979). The dimension of the covariate space was taken to be of two and three dimensions. A single data set was generated for each combination of the parameters

$$n = 20, 50,$$

$$\kappa = 1, 3, 5,$$

$$\beta_i = .1, 2.1,$$

from the Fisher and Lee model. The x_i 's were taken as $U(-1, 1)$. For each data set the algorithm was run three hundred (300) times with different seeds creating different original stores. For each data set two store sizes were also studied. Based on some preliminary studies, doubling the size of the store recommended by Conlon worked well. Thus the store sizes tested were $20 \times (\dim \mathbf{x}_i)$ and $40 \times (\dim \mathbf{x}_i)$. The following tables summarize the results.

Tables B.1 and B.2 show the results of the simulation study. For Table B.2, the case marked by the asterisk(*), represents a situation in which the log-likelihood components of two distinct points achieved maxima that were indistinguishable from one another to three decimal places. Basically both models fit equally well and the algorithm always converged to one or the other maxima.

The results of Tables B.1 and B.2 imply that the modified version of Price's algorithm worked well for large κ and n . For samples of size 50 the modification of Price's algorithm was able to find the true global maximum on the likelihood

Table B.1. Simulation results of Price's Algorithm: two dimensional case.

n=20		STORE 40		STORE 80	
	Conc.	% Worked	Good/Total	% Worked	Good/Total
$\beta_1 = .1 \quad \beta_2 = .1$	$\kappa = 1$	60%	905/1540	88%	1645/2859
	$\kappa = 3$	100%	445/827	100%	855/1585
	$\kappa = 5$	100%	587/1079	100%	1090/2028
$\beta_1 = 2.1 \quad \beta_2 = .1$	$\kappa = 1$	100%	598/1293	100%	1197/2556
	$\kappa = 3$	100%	619/1283	100%	1241/2568
	$\kappa = 5$	100%	630/1316	100%	1253/2598
$\beta_1 = 2.1 \quad \beta_2 = 2.1$	$\kappa = 1$	100%	576/1260	100%	1152/2503
	$\kappa = 3$	100%	626/1275	100%	1248/2521
n=50					
$\beta_1 = .1 \quad \beta_2 = .1$	$\kappa = 1$	98%	706/1308	99%	1279/2383
	$\kappa = 3$	100%	586/1085	100%	1124/2099
	$\kappa = 5$	100%	668/1241	100%	1256/2349
$\beta_1 = 2.1 \quad \beta_2 = .1$	$\kappa = 1$	100%	656/1347	100%	1308/2686
	$\kappa = 3$	100%	883/1391	100%	1368/2754
	$\kappa = 5$	100%	680/1387	100%	1363/2759
$\beta_1 = 2.1 \quad \beta_2 = 2.1$	$\kappa = 1$	100%	667/1315	100%	1270/2563
	$\kappa = 3$	100%	662/1333	100%	1324/2660
	$\kappa = 5$	100%	672/1370	100%	1354/2736

surface more than ninety-eight percent of the time. For small concentrations and small sample sizes the algorithm does not guarantee finding the global maximum of the likelihood surface. The Good/Total values in the tables represent the average number of function evaluations that resulted in improvements and the average number of function evaluations required for the algorithm to converge. It is suggested that the algorithm be run several times to insure that the global maximum is actually found. The behavior of the algorithm improves as either the sample size gets larger, and/or the concentration of the underlying distribution increases. The algorithm as defined has a number of benefits over other algorithms. For example, Price's algorithm, and its modification, require only evaluation of the likelihood, and hence is not computationally intensive. Also based on these simulation results and others

like them, the size of the store seems to increase only linearly with the dimensionality of the problem.

Table B.2. Simulation results of Price's Algorithm: three dimensional case.

n=20			STORE 40	STORE 80
	Conc.	% Worked	Good/Total	% Worked Good/Total
$\beta_1 = .1 \quad \beta_2 = .1$ $\beta_3 = .1 \quad \beta_4 = .1$	$\kappa = 1$	87%	2163/6319	97% 4205/13139
	$\kappa = 3$	100%	2507/4575	100% 3830/7197
	$\kappa = 5$	100%	1213/2397	100% 2348/4658
$\beta_1 = 2.1 \quad \beta_2 = .1$ $\beta_3 = .1 \quad \beta_4 = .1$	$\kappa = 1$	85%	2102/7761	98% 4244/16280
	$\kappa = 3$	100%	2111/4873	100% 4221/9700
	$\kappa = 5$	98%	2131/5794	100% 4244/11622
$\beta_1 = 2.1 \quad \beta_2 = 2.1$ $\beta_3 = .1 \quad \beta_4 = .1$	$\kappa = 1$	92%	2057/6959	98% 4158/14015
	$\kappa = 3$	93%	2189/6085	99% 4397/12187
	$\kappa = 5$	100%	2124/4889	100% 4245/9723
$\beta_1 = 2.1 \quad \beta_2 = 2.1$ $\beta_3 = 2.1 \quad \beta_4 = .1$	$\kappa = 1$	100%	2081/6744	100% 4175/13479
	$\kappa = 3$	100%	2095/4592	100% 4186/9141
	$\kappa = 5$	100%	2031/4503	100% 4065/8971
$\beta_1 = 2.1 \quad \beta_2 = 2.1$ $\beta_3 = 2.1 \quad \beta_4 = 2.1$	$\kappa = 1$	79%	2417/6575	94% 4754/17615
	$\kappa = 3$	72%	2237/7971	94% 4754/17615
	$\kappa = 5$	67%	2287/8252	84% 4549/17044
n=50				
$\beta_1 = .1 \quad \beta_2 = .1$ $\beta_3 = .1 \quad \beta_4 = .1$	$\kappa = 1$	99%	2360/4359	100% 3784/7163
	$\kappa = 3$	100%	4713/8408	100% 6089/11266
	$\kappa = 5$	100%	3220/5886	100% 4606/8729
$\beta_1 = 2.1 \quad \beta_2 = .1$ $\beta_3 = .1 \quad \beta_4 = .1$	$\kappa = 1$	100%	2194/5287	100% 4411/10615
	$\kappa = 3$	100%	2262/4953	100% 4534/9910
	$\kappa = 5$	100%	2297/5039	100% 4600/10095
$\beta_1 = 2.1 \quad \beta_2 = 2.1$ $\beta_3 = .1 \quad \beta_4 = .1$	$\kappa = 1$	100%	2095/5571	100% 4224/11147
	$\kappa = 3$	100%	2226/4922	100% 4469/9796
	$\kappa = 5$	100%	2225/4820	100% 4471/9756
$\beta_1 = 2.1 \quad \beta_2 = 2.1$ $\beta_3 = 2.1 \quad \beta_4 = .1$	$\kappa = 1$	*42%*	2060/5711	*48%* 4215/11814
	$\kappa = 3$	100%	2102/4575	100% 4223/9156
	$\kappa = 5$	100%	2141/4648	100% 4280/9250
$\beta_1 = 2.1 \quad \beta_2 = 2.1$ $\beta_3 = 2.1 \quad \beta_4 = 2.1$	$\kappa = 1$	100%	2096/5498	100% 4201/10913
	$\kappa = 3$	100%	2138/4663	100% 4287/9315
	$\kappa = 5$	100%	2156/4682	100% 4330/9384

APPENDIX C ESTIMATING THE LENGTH OF A VECTOR GIVEN THE DIRECTION

Since the observations are assumed to be independent, calculating $E(L_i|\Theta, \beta_1, \beta_2)$ requires conditioning only on the observed angle and not on the entire vector of directions. Assuming β_1 and β_2 known is equivalent to assuming the mean vector of the bivariate normal is known and this implies that κ and μ are known. Thus, it is sufficient to find the expectation for an arbitrary known κ and known μ . By assuming

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} L \cos(\theta) \\ L \sin(\theta) \end{pmatrix} \sim BN \left(\begin{pmatrix} \kappa \cos(\mu) \\ \kappa \sin(\mu) \end{pmatrix}, I \right)$$

and letting $Q = \kappa \cos(\theta - \mu)$, it follows that

$$f(L|\theta, \kappa, \mu) = \frac{L \exp \left(-\frac{1}{2}L^2 - LQ \right)}{1 + \sqrt{2\pi}Q \exp \left(\frac{1}{2}Q^2 \right) \Phi(Q)}.$$

Hence

$$\begin{aligned} E(L|\theta, \kappa, \mu) &= \frac{1}{1 + \sqrt{2\pi}Q \exp \left(\frac{1}{2}Q^2 \right) \Phi(Q)} \int_0^\infty L^2 \exp \left(-\frac{L^2 - 2LQ}{2} \right) dL \\ &= \frac{\exp \left(\frac{1}{2}Q^2 \right)}{1 + \sqrt{2\pi}Q \exp \left(\frac{1}{2}Q^2 \right) \Phi(Q)} \int_0^\infty L^2 \exp \left(-\frac{(L - Q)^2}{2} \right) dL. \end{aligned}$$

Now by completing the square from L^2 to $(L - Q)^2$ within the integral and integrating each term, the first term by parts, gives

$$\begin{aligned}
\mathbf{E}(L|\theta, \kappa, \mu) &= \frac{\exp\left(\frac{1}{2}Q^2\right)}{1 + \sqrt{2\pi}Q \exp\left(\frac{1}{2}Q^2\right) \Phi(Q)} \left[Q \exp\left(-\frac{1}{2}Q^2\right) + \right. \\
&\quad \left. \sqrt{2\pi}\Phi(Q) + \sqrt{2\pi}Q^2\Phi(Q) \right] \\
&= Q + \frac{\sqrt{2\pi} \exp\left(\frac{1}{2}Q^2\right) \Phi(Q)}{1 + \sqrt{2\pi}\hat{Q} \exp\left(\frac{1}{2}Q^2\right) \Phi(Q)},
\end{aligned}$$

which is equivalent to the expression given in equation (4.12).

APPENDIX D NON-NEGATIVE DEFINITENESS OF THE HESSIAN

The negative of the Hessian matrix for the combined vector $(\beta'_1 : \beta'_2)$ is found in the text through equations 4.17, 4.18 and 4.19. A sufficient condition for the negative of the Hessian matrix to be non-negative definite is for the matrix \mathbf{V} to be non-negative definite. A necessary and sufficient condition for this is that all principle minors of \mathbf{V} are greater than or equal to zero. Each principle minor can be written in the form

$$\prod_i \{1 - w_{ii} \cos^2(\theta_i)\} \prod_j \{1 - w_{jj} \sin^2(\theta_j)\} \prod_k (1 - w_{kk}),$$

where each of the products are over mutually exclusive sets. Since

$$0 \leq \sin^2(\theta) \leq 1 \quad \text{and} \quad 0 \leq \cos^2(\theta) \leq 1,$$

it is sufficient to verify that $(1 - w_{ii})$ is positive for all i .

Equation 4.19 implies

$$1 - w_{ii} = \frac{1}{\left\{1 + \sqrt{2\pi} Q_i \Phi(Q_i) \exp\left(\frac{1}{2} Q_i^2\right)\right\}^2} \left[2\pi \Phi^2(Q_i) \exp(Q_i^2) - 1 - \sqrt{2\pi} Q_i \Phi(Q_i) \exp\left(\frac{1}{2} Q_i^2\right) \right]. \quad (\text{D.1})$$

The first term on the right hand side is nonnegative since it is a squared term. The second term on the right can be written as

$$\frac{1}{\phi^2 Q_i} \left[\Phi^2(Q_i) - \phi^2(Q_i) - Q_i \phi(Q_i) \Phi(Q_i) \right],$$

where $\phi(\cdot)$ is the PDF of the standard normal. This term is greater than or equal to zero if and only if

$$\Phi^2(Q_i) - \phi^2(Q_i) - Q_i\phi(Q_i)\Phi(Q_i) \quad (\text{D.2})$$

is greater than or equal to zero. To prove this last statement, it is sufficient to show that equation D.2 is a strictly increasing function of Q_i and that the limit as $Q_i \rightarrow -\infty$ is zero. These two properties together prove that equation (D.2) is strictly positive and hence the right hand side of equation (D.1) is strictly positive.

The derivative of equation (D.2) with respect to Q_i is

$$\Phi(Q_i)\phi(Q_i) + Q_i\phi^2(Q_i) + Q_i^2\Phi(Q_i)\phi(Q_i). \quad (\text{D.3})$$

It is obvious that this term is positive for Q_i positive. Thus, only the case when Q_i is negative need be proved. By an application of Mill's Ratio (See Resnick (1992)) it follows that for Q_i negative

$$\Phi(Q_i) > \frac{-Q_i\phi(Q_i)}{1 + Q_i^2}.$$

By using this in equation (D.3) it can be seen that

$$\Phi(Q_i)\phi(Q_i) + Q_i\phi^2(Q_i) + Q_i^2\Phi(Q_i)\phi(Q_i) > 0.$$

Thus, equation (D.2) is an increasing function of Q_i .

Next it is easy to verify that

$$\lim_{Q_i \rightarrow -\infty} [\Phi^2(Q_i) - \phi^2(Q_i) - Q_i\phi(Q_i)\Phi(Q_i)] = 0. \quad (\text{D.4})$$

The first two terms in equation (D.4) are obviously 0 in the limit. The last term in equation (D.4) is less than $Q_i/\exp(\frac{1}{2}Q_i^2)$, which goes to zero in the limit. Hence, equation (D.2) is a positive function, which implies that the right hand side of equation (D.1) is also positive. Hence each term $1 - w_{ii}$ is nonnegative, indicating that each principle minor is nonnegative, which in turn indicates that the Hessian matrix is non-negative definite.

REFERENCES

- Abramowitz, M. and Stegun, I. (1970). *Handbook of Mathematical Functions*, National Bureau of Standards, New York.
- Agresti, A. (1990). *Categorical Data Analysis*, John Wiley & Sons, New York.
- Anderson, C. M. and Wu, C. F. J. (1994a). Dispersion measures and analysis of factorial directional data with replicates, *Technical report*, Institute for the Improvement of Quality and Productivity, Dept. of Stat. and Act. Sci., Univ. of Waterloo.
- Anderson, C. M. and Wu, C. F. J. (1994b). Measuring location effects from factorial experiments with a directional response, *Technical report*, Department of Statistical and Actuarial Sciences, Univ. of Western Ontario.
- Becker, R. A., Chambers, J. A. and Wilks, A. R. (1988). *The New S Language*, Wadsworth & Brooks, Pacific Grove, California.
- Best, D. J. and Fisher, N. I. (1979). Efficient simulation of the von Mises distribution, *Applied Statistics* **24**: 152–157.
- Best, D. J. and Fisher, N. I. (1981). The bias of the maximum likelihood estimators of the von Mises-Fisher concentration parameters, *Communications in Statistics, Part B-Simulation and Computation* **10**: 493–502.
- Bohachevsky, I. O., Johnson, M. E. and Stein, M. L. (1986). Generalized simulated annealing for function optimization, *Technometrics* **28**: 209–217.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association* **74**: 829–836.
- Conlon, M. (1992). The controlled random search procedure for function optimization, *Communications in Statistics, Part B-Simulation and Computation* **21**: 919–923.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society, Series B, Methodological* **39**: 1–38.
- Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information, *Biometrika* **65**: 457–481.
- Fisher, N. (1993). *Statistical Analysis of Circular Data*, Cambridge Univ. Press, Cambridge.


- Fisher, N. I. and Lee, A. J. (1992). Regression models for an angular response, *Biometrics* **48**: 665-677.
- Fisher, R. A. (1973). *Statistical Methods and Scientific Inference*, 3rd Edition, Oliver and Boyd, Edinburgh.
- Gould, A. L. (1969). A regression technique for angular variates, *Biometrics* **25**: 683-700.
- Harrison, D. and Kanji, G. K. (1988). The development of analysis of variance for circular data, *Journal of Applied Statistics* **15**: 197-223.
- Harrison, D., Kanji, G. K. and Gadsen, R. J. (1986). Analysis of variance for circular data, *Journal of Applied Statistics* **13**: 123-138.
- Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case, *The Annals of Mathematical Statistics* **42**: 1977-1991.
- Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models, *Journal of the American Statistical Association* **73**: 602-606.
- Klotz, J. (1964). Small-sample power of the bivariate sign tests of Blumen and Hodges, *The Annals of Mathematical Statistics* **35**: 1576-1582.
- Mardia, K. (1972). *Statistics of Directional Data*, Academic Press, New York.
- Marsaglia, G. and Tsang, W. W. (1984). A fast, easily implemented method for sampling from decreasing or symmetric unimodal density functions, *SIAM Journal on Scientific and Statistical Computing* **5**: 349-359.
- Price, W. L. (1977). Optimization, *The Computer Journal* **20**: 367-370.
- Resnick, S. I. (1992). *Adventures in Stochastic Processes*, Birkhaeuser, Boston.
- Stephens, M. (1969). Tests for the von Mises distribution, *Biometrika* **56**: 149-160.
- Stephens, M. (1982). Use of the von Mises distribution to analyze continuous proportions, *Biometrika* **62**: 197-203.
- Underwood, A. and Chapman, M. (1985). Multifactorial analyses of directions of movement of animals, *J. Exp. Mar. Biol. Ecol.* **91**: 17-43.
- Underwood, A. and Chapman, M. (1989). Experimental analyses of the influences of topography of the substratum on movements and density of an intertidal snail, *littorina unisfasciata*, *J. Exp. Mar. Biol. Ecol.* **134**: 175-196.
- Wald, A. (1949). Note on the consistency of the maximum likelihood estimate, *The Annals of Mathematical Statistics* **20**: 595-601.
- Walker, T. J. and Littell, R. C. (1994). Orientation of fall-migrating butterflies in North Peninsular Florida and source areas, *Ethology* **98**: 60-84.
- Watson, G. and Williams, E. J. (1956). On the construction of significance tests on the circle and the sphere, *Biometrika* **43**: 344-352.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm, *The American Statistician* **11**: 95-103.

BIOGRAPHICAL SKETCH

Scott Porter Morrison was born on September 30, 1968, in Danville, Virginia. Soon thereafter he and his family moved to Venice, Florida. After 17 years in Florida without seeing snow, he left to attend Wake Forest University, where he graduated with a B.S. in mathematics along with a minor in psychology in 1990. Instead of pursuing a career in the charter fishing industry, Scott decided to attend graduate school at the University of Florida. Scott received a Master of Statistics degree in 1992, and will receive a Ph.D. degree in statistics in December of 1995. While at the University of Florida, Scott was a teaching assistant for two years, and worked as a research assistant with Dr. Ramon Littell for three plus years. Outside of school, Scott is a diehard fisherman, spending every summer since 1989 working in the Tarpon fishing industry in Boca Grande, Florida. Besides fishing, Scott enjoys most sports, particularly basketball and golf. After graduation, Scott, his wife Alicia, and dog Cayman are moving to Houston, Texas, where Scott has accepted a position with Shell Development Company.

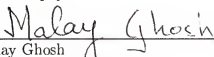
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.


Ramon C. Littell, Chairman
Professor of Statistics

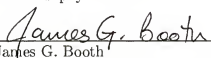
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.


Brett D. Presnell
Assistant Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.


Malay Ghosh
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.


James G. Booth
Associate Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.


Thomas J. Walker
Professor of Entomology and Nematology

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 1995

Dean, Graduate School